

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 12:55:49 ; Search time 91.9545 Seconds  
(without alignments)  
54.616 Million cell updates/sec

Title: US-10-067-620-1  
Perfect score: 64  
Sequence: 1 XXSGISNTVYANPK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query | Length | DB | ID       | Description        |
|------------|-------|---------|--------|----|----------|--------------------|
|            |       | Match   |        |    |          |                    |
| 1          | 62    | 96.9    | 14     | 5  | ABB81968 | Abb81968 30 kDa ra |
| 2          | 39    | 60.9    | 106    | 2  | AAW88497 | Aaw88497 Human epi |
| 3          | 39    | 60.9    | 106    | 3  | AAG03962 | Aag03962 Human sec |
| 4          | 39    | 60.9    | 120    | 8  | ADJ37083 | Adj37083 Human per |
| 5          | 39    | 60.9    | 329    | 7  | ADC94153 | Adc94153 E. faeciu |
| 6          | 39    | 60.9    | 401    | 4  | ABB06884 | Abb06884 Micromono |
| 7          | 39    | 60.9    | 474    | 4  | AAU04831 | Aau04831 Micromono |
| 8          | 39    | 60.9    | 489    | 6  | ABP99325 | Abp99325 Orthosomy |
| 9          | 39    | 60.9    | 490    | 6  | ABU31218 | Abu31218 Protein e |

|    |    |      |      |   |          |          |            |
|----|----|------|------|---|----------|----------|------------|
| 10 | 39 | 60.9 | 493  | 6 | ABP99326 | Abp99326 | Orthosomy  |
| 11 | 39 | 60.9 | 637  | 7 | ABO62726 | Abo62726 | Klebsiell  |
| 12 | 39 | 60.9 | 858  | 2 | AAR30477 | Aar30477 | Human leu  |
| 13 | 39 | 60.9 | 1024 | 2 | AAW19604 | Aaw19604 | Mycoplasma |
| 14 | 38 | 59.4 | 120  | 4 | AAG65232 | Aag65232 | Human RNA  |
| 15 | 38 | 59.4 | 329  | 6 | ABO00477 | Abo00477 | Novel hum  |
| 16 | 38 | 59.4 | 336  | 6 | ABU49299 | Abu49299 | Protein e  |
| 17 | 38 | 59.4 | 432  | 3 | AAB18414 | Aab18414 | Amino aci  |
| 18 | 38 | 59.4 | 432  | 4 | AAB67570 | Aab67570 | Amino aci  |
| 19 | 38 | 59.4 | 433  | 4 | AAM40128 | Aam40128 | Human pol  |
| 20 | 38 | 59.4 | 440  | 3 | AAB43115 | Aab43115 | Human ORF  |
| 21 | 38 | 59.4 | 505  | 4 | AAM41914 | Aam41914 | Human pol  |
| 22 | 38 | 59.4 | 602  | 4 | ABG02902 | Abg02902 | Novel hum  |
| 23 | 38 | 59.4 | 1851 | 4 | AAU00023 | Aau00023 | Human act  |
| 24 | 38 | 59.4 | 1851 | 5 | AAO17860 | Aao17860 | Pyrin dom  |
| 25 | 37 | 57.8 | 138  | 5 | ABP11408 | Abp11408 | Human ORF  |
| 26 | 37 | 57.8 | 219  | 4 | AAB63630 | Aab63630 | Human gas  |
| 27 | 37 | 57.8 | 268  | 4 | AAB63629 | Aab63629 | Human gas  |
| 28 | 37 | 57.8 | 376  | 6 | ABG72091 | Abg72091 | Piromyces  |
| 29 | 37 | 57.8 | 389  | 3 | AAV75076 | Aay75076 | Neisseria  |
| 30 | 37 | 57.8 | 389  | 3 | AAV75077 | Aay75077 | Neisseria  |
| 31 | 37 | 57.8 | 389  | 3 | AAV75078 | Aay75078 | Neisseria  |
| 32 | 37 | 57.8 | 399  | 4 | ABB65138 | Abb65138 | Drosophil  |
| 33 | 37 | 57.8 | 441  | 6 | ABP78454 | Abp78454 | N. gonorr  |
| 34 | 37 | 57.8 | 515  | 7 | ADC95690 | Adc95690 | E. faeciu  |
| 35 | 37 | 57.8 | 586  | 6 | ABJ26507 | Abj26507 | Aspergill  |
| 36 | 37 | 57.8 | 586  | 6 | ABJ25907 | Abj25907 | Aspergill  |
| 37 | 37 | 57.8 | 700  | 4 | AAG83279 | Aag83279 | Chlamydia  |
| 38 | 37 | 57.8 | 700  | 5 | ABB94250 | Abb94250 | Chlamydia  |
| 39 | 37 | 57.8 | 712  | 2 | AAR06471 | Aar06471 | Derived p  |
| 40 | 37 | 57.8 | 712  | 2 | AAV33359 | Aay33359 | Human isl  |
| 41 | 37 | 57.8 | 712  | 2 | AAV49853 | Aay49853 | Human pan  |
| 42 | 37 | 57.8 | 712  | 2 | AAW80485 | Aaw80485 | Islet cel  |
| 43 | 37 | 57.8 | 712  | 3 | AAV49327 | Aay49327 | Pancreati  |
| 44 | 37 | 57.8 | 731  | 4 | AAB48573 | Aab48573 | Human bre  |
| 45 | 37 | 57.8 | 829  | 6 | AAO29907 | Aao29907 | Human org  |

# ALIGNMENTS

## RESULT 1

ABB81968

ID ABB81968 standard; peptide; 14 AA.

XX

AC ABB81968;

XX

DT 25-NOV-2002 (first entry)

XX

DE 30 kDa ragweed pollen allergen tryptic peptide 1.

XX

KW Ragweed; pollen; allergen; Ambt 7; glycoprotein; antiallergic;

KW immunotherapy; disulphide protein.

XX

OS Ambrosia elatior.

XX

FH Key Location/Qualifiers

FT Misc-difference 1  
 FT /label= Leu or Ile  
 FT Misc-difference 2  
 FT /label= Leu or Ile  
 XX  
 PN WO200263012-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 04-FEB-2002; 2002WO-US003346.  
 XX  
 PR 05-FEB-2001; 2001US-0266686P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Buchanan BB, Del Val G, Frick OL;  
 XX  
 DR WPI; 2002-657539/70.  
 XX  
 PT New ragweed pollen allergens, useful in allergy testing and immunotherapy,  
 PT regimens, particularly for treating sensitivity to pollen or pollen  
 PT allergy (e.g. anaphylaxis, or symptoms of hives or asthma) in a mammal,  
 PT especially a human.  
 XX  
 PS Claim 1; Page 53; 70pp; English.  
 XX  
 CC The invention relates to an isolated pollen allergen purified from  
 CC ragweed pollen, substantially free of any other pollen proteins, or a  
 CC protein that is an antigenic fragment of a pollen allergen Ambt 7. The  
 CC allergen is characterized by the following physiochemical and biological  
 CC properties: (a) being contained in pollen extracts; (b) a glycoprotein;  
 CC (c) a sulphydryl group containing protein; (d) a molecular weight of  
 CC about 30 kDa as determined by SDS-polyacrylamide gel electrophoresis; and  
 CC (e) possessing allergen activity. The pollen allergen, or antigenic  
 CC protein fragment of the pollen allergen Ambt 7, or composition is useful  
 CC for treating sensitivity to pollen or pollen allergy in a mammal. This  
 CC allergy includes anaphylaxis or atopy, which includes the symptoms of hay  
 CC fever, asthma or hives. The allergen is also useful in allergy testing  
 CC and immunotherapy regimens. Sequences ABB81968-978 represent tryptic  
 CC peptide fragments of the 30 kDa ragweed complete pollen extract  
 CC disulphide protein allergen  
 XX  
 SQ Sequence 14 AA;

Query Match 96.9%; Score 62; DB 5; Length 14;  
 Best Local Similarity 100.0%; Pred. No. 0.00013;  
 Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
 |||||  
 Db 3 SGISNTVYANPK 14

RESULT 2  
 AAW88497  
 ID AAW88497 standard; protein; 106 AA.  
 XX

AC AAW88497;  
 XX  
 DT 30-MAR-1999 (first entry)  
 XX  
 DE Human epidermoid carcinoma clone HP10389-encoded protein.  
 XX  
 KW Transmembrane protein; HP10389; human; epidermoid carcinoma.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9855508-A2.  
 XX  
 PD 10-DEC-1998.  
 XX  
 PF 03-JUN-1998; 98WO-JP002445.  
 XX  
 PR 03-JUN-1997; 97JP-00144948.  
 XX  
 PA (SAGA ) SAGAMI CHEM RES CENTRE.  
 PA (PROT-) PROTEGENE INC.  
 XX  
 PI Kato S, Sekine S, Yamaguchi T;  
 XX  
 DR WPI; 1999-045730/04.  
 DR N-PSDB; AAV84365.  
 XX  
 PT New human proteins containing transmembrane domains and their encoding  
 PT sequences - useful in the preparation of antibodies and large-scale  
 PT protein production, gene diagnosis, and gene therapy.  
 XX  
 PS Claim 1; Page 133-134; 178pp; English.  
 XX  
 CC This is the amino acid sequence of a transmembrane protein encoded by  
 CC human epidermoid cancer cDNA clone HP10389 (see AAV84365). The encoded  
 CC protein has 2 putative transmembrane domains. It shows no homology to  
 CC protein database sequences. The invention provides nucleotide sequences  
 CC (see AAV84359-76) coding for 18 transmembrane proteins (see AAW88491-  
 CC 508), vectors containing such polynucleotides, and eukaryotic cells  
 CC containing the vectors. The proteins can be used as antigens or as  
 CC compositions in the preparation of antibodies against the proteins. The  
 CC polynucleotides can be used as probes for gene diagnosis, and as gene  
 CC sources for gene therapy and large-scale production of proteins encoded  
 CC by the cDNA. The host cells are used for the detection of ligands  
 CC corresponding to the expressed proteins, and the screening of low mol.wt.  
 CC medicines  
 XX  
 SQ Sequence 106 AA;

Query Match 60.9%; Score 39; DB 2; Length 106;  
 Best Local Similarity 63.6%; Pred. No. 28;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14  
 |:| ||| |:|  
 Db 23 GLSPTVYRNPE 33



RESULT 3

AAG03962

ID AAG03962 standard; protein; 106 AA.

XX

AC AAG03962;

XX

DT 06-OCT-2000 (first entry)

XX

DE Human secreted protein, SEQ ID NO: 8043.

XX

KW Human; 5' EST; expressed sequence tag; secreted protein; cDNA isolation;

KW gene therapy; chromosome mapping.

XX

OS Homo sapiens.

XX

PN EP1033401-A2.

XX

PD 06-SEP-2000.

XX

PF 21-FEB-2000; 2000EP-00200610.

XX

PR 26-FEB-1999; 99US-0122487P.

XX

PA (GEST ) GENSET.

XX

PI Dumas Milne Edwards J, Duclert A, Giordano J;

XX

DR WPI; 2000-500381/45.

DR N-PSDB; AAC03968.

XX

PT New nucleic acid that is a 5' expressed sequence tag (5' EST) for  
PT obtaining cDNAs and genomic DNAs that correspond to 5'ESTs and for  
PT diagnostic, forensic, gene therapy and chromosome mapping procedures.

XX

PS Claim 13; SEQ ID NO 8043; 71pp + Sequence Listing; English.

XX

CC The present sequence is a polypeptide encoded by one of a large number of  
CC 5' ESTs derived from mRNAs encoding secreted proteins. The 5' ESTs were  
CC prepared from total human RNAs or polyA+ RNAs derived from 30 different  
CC tissues. EST sequences usually correspond mainly to the 3' untranslated  
CC region (UTR) of the mRNA because they are often obtained from oligo-dT  
CC primed cDNA libraries. Such ESTs are not well suited for isolating cDNA  
CC sequences derived from the 5' ends of mRNAs and even in those cases where  
CC longer cDNA sequences have been obtained, the full 5' UTR is rarely  
CC included. 5' ESTs are derived from mRNAs with intact 5' ends and can  
CC therefore be used to obtain full length cDNAs and genomic DNAs. 5' ESTs  
CC are also used in diagnostic, forensic, gene therapy and chromosome  
CC mapping procedures. They are used to obtain upstream regulatory sequences  
CC and to design expression and secretion vectors

XX

SQ Sequence 106 AA;

Query Match 60.9%; Score 39; DB 3; Length 106;

Best Local Similarity 63.6%; Pred. No. 28;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy

4 GISNTVYANPK 14

Db                   |:| ||| |:|  
23 GLSPTVYRNPE 33

RESULT 4

ADJ37083

ID   ADJ37083 standard; protein; 120 AA.

XX

AC   ADJ37083;

XX

DT   06-MAY-2004   (first entry)

XX

DE   Human pertubagen SEQ ID NO:8.

XX

KW   cytotoxicity; conditional cytotoxicity; cell-specific cytotoxicity;  
KW   pertubagen; human.

XX

OS   Homo sapiens.

XX

PN   WO2004012574-A2.

XX

PD   12-FEB-2004.

XX

PF   16-JUL-2003; 2003WO-US022241.

XX

PR   16-JUL-2002; 2002US-0396171P.

XX

PA   (DELT-) DELTAGEN PROTEOMICS INC.

XX

PI   Kamb CA;

XX

DR   WPI; 2004-156982/15.

XX

PT   Inducing cytotoxicity in a cell comprises contacting the cell with a  
PT   peptide fragment, or introducing into a cell a polynucleotide encoding  
PT   the peptide fragment, effective to induce cytotoxicity.

XX

PS   Claim 14; SEQ ID NO 8; 78pp; English.

XX

CC   The invention relates to a novel method for inducing cytotoxicity in a  
CC   cell comprises contacting the cell with a peptide fragment, or  
CC   introducing into a cell a polynucleotide encoding the peptide fragment,  
CC   effective to induce cytotoxicity. The method of the invention is useful  
CC   for evaluation of conditional cytotoxicity and cell-specific  
CC   cytotoxicity. The present sequence represents a protein of the invention.

XX

SQ   Sequence 120 AA;

Query Match                   60.9%;   Score 39;   DB 8;   Length 120;  
Best Local Similarity   63.6%;   Pred. No. 33;  
Matches       7;   Conservative       2;   Mismatches       2;   Indels       0;   Gaps       0;

Qy                   4 GISNTVYANPK 14

|:| ||| |:|

Db                   37 GLSPTVYRNPE 47

RESULT 5

ADC94153

ID ADC94153 standard; protein; 329 AA.

XX

AC ADC94153;

XX

DT 01-JAN-2004 (first entry)

XX

DE E. faecium protein sequence SEQ ID 3780.

XX

KW Vaccine; urinary tract infection; bacteraemia; endocarditis; wound;  
KW abdominal-pelvic infection.

XX

OS Enterococcus faecium.

XX

PN US6583275-B1.

XX

PD 24-JUN-2003.

XX

PF 30-JUN-1998; 98US-00107532.

XX

PR 02-JUL-1997; 97US-0051571P.

PR 14-MAY-1998; 98US-0085598P.

XX

PA (GENO-) GENOME THERAPEUTICS CORP.

XX

PI Doucette-Stamm LA, Bush D;

XX

DR WPI; 2003-799836/75.

DR N-PSDB; ADC90499.

XX

PT New isolated nucleic acid derived from Enterococcus faecium encoding an  
PT Enterococcus faecium polypeptide useful for detection, prevention and  
PT treatment of a pathological condition resulting from a bacterial  
PT infection.

XX

PS Example 1; SEQ ID NO 3780; 243pp; English.

XX

CC The invention relates to an isolated nucleic acid derived from  
CC Enterococcus faecium encoding an Enterococcus faecium polypeptide having  
CC one of 10 fully defined sequences given in the (or comprising 40  
CC sequential nucleotides chosen from any of the nucleic acids, its  
CC complement or sequences hybridising to it). Also included are a  
CC recombinant vector comprising the nucleic acid operably linked to  
CC transcription regulatory element, a cell comprising the vector and a  
CC single-stranded probe comprising the nucleic acid. The nucleic acids are  
CC chosen from 3654 disclosed sequences encoding 3654 disclosed proteins.  
CC The nucleic acids is useful for diagnosing pathological conditions  
CC resulting from E. faecium bacterial infection (e.g. urinary tract  
CC infection, bacteraemia, endocarditis, wounds and abdominal-pelvic  
CC infection) and for screening drugs such as agonists and antagonists. The  
CC nucleic acid is useful for recombinant production of Candida albicans -  
CC derived peptides or antisense polypeptides. Pharmaceutical compositions  
CC and vaccines containing the nucleic acid are useful for preventing or  
CC treating Enterococcus faecium infections. The present sequence represents  
CC one if the disclosed E. faecium proteins.

XX

SQ Sequence 329 AA;

Query Match 60.9%; Score 39; DB 7; Length 329;  
Best Local Similarity 60.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

QY 3 SGISNTVYAN 12  
||:|:|:|:  
Db 162 SGVSNSIHAN 171

RESULT 6

ABB06884

ID ABB06884 standard; protein; 401 AA.

XX

AC ABB06884;

XX

DT 18-JUN-2002 (first entry)

XX

DE Micromonospora carbonacea everninomicin locus protein ORF 4.

XX

KW Micromonospora carbonacea; antibiotic; everninomicin; biosynthesis;  
KW gene cluster; genetic manipulation; contig.

XX

OS Micromonospora carbonacea.

XX

PN WO200155180-A2.

XX

PD 02-AUG-2001.

XX

PF 29-JAN-2001; 2001WO-CA000128.

XX

PR 27-JAN-2000; 2000US-0177711P.

XX

PA (ECOP-) ECOPIA BIOSCIENCES INC.

PA (FARN/) FARNET C.

XX

PI Staffa A, Zazopoulos E, Mercure S, Nowacki P;

XX

DR WPI; 2001-476185/51.

DR N-PSDB; ABL50557.

XX

PT Novel isolated gene cluster encoding polypeptides involved in  
PT everninomicin biosynthesis useful for construction of everninomicin  
PT overproducing strains, and to allow chemical modifications of  
PT everninomicin to enhance certain properties.

XX

PS Claim 15; Fig 1; 181pp; English.

XX

CC ABL50555 to ABL50562 represent contigs 1 to 8 from the Micromonospora  
CC carbonacea everninomicin biosynthetic locus gene cluster. The contigs  
CC encode the protein sequences designated ORF (open reading frame) 1 to 49,  
CC given in ABB06881 to ABB06930. The gene cluster is useful for the  
CC construction of the everninomicin antibiotic in overproducing strains,  
CC and to allow chemical modifications of everninomicin to enhance certain  
CC properties via genetic manipulation or combinational biosynthesis. The  
CC gene cluster can be used to produce genetic systems and genes encoding

CC novel enzyme activities, and avoid the problems of low yield and quality  
CC of everninomicins produced by chemical synthesis  
XX  
SQ Sequence 401 AA;

Query Match 60.9%; Score 39; DB 4; Length 401;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
:|: :||| :||  
Db 351 TGLKDTVYVSPK 362

RESULT 7

AAU04831

ID AAU04831 standard; protein; 474 AA.

XX

AC AAU04831;

XX

DT 11-SEP-2003 (revised)

DT 26-SEP-2001 (first entry)

XX

DE Micromonospora everninomicin biosynthetic enzyme evrG.

XX

KW Everninomicin; antibiotic; bottle-neck gene; orthomycin; fermentation;  
KW Tailoring gene product; oxidase; evrG.

XX

OS Micromonospora sp. ATCC 39149.

XX

PN WO200151639-A2.

XX

PD 19-JUL-2001.

XX

PF 12-JAN-2001; 2001WO-US001187.

XX

PR 12-JAN-2000; 2000US-0175751P.

XX

PA (SCHE ) SCHERING CORP.

XX

PI Hosted TJ, Horan AC, Wang TX;

XX

DR WPI; 2001-442147/47.

DR N-PSDB; AAS08693.

XX

PT New nucleic acid molecules encoding everninomicin pathway gene products,  
PT useful for improving yields of everninomicin, to produce new  
PT everninomicin and as probes to identify homologous sequences.

XX

PS Claim 19; Fig 11; 109pp; English.

XX

CC The sequence represents a Tailoring gene product, an oxidase, evrG. The  
CC protein comprises one of 98 enzymes of the everninomicin antibiotic  
CC biosynthetic pathway. A vector comprising a M. carbonacea everninomicin  
CC biosynthetic pathway resistance gene product is useful for selecting for  
CC a transfected or transformed host cell. An integrative version of the  
CC vector is useful for introducing a everninomicin pathway gene (a bottle-

CC neck gene) into an actinomycete of the genus *Micromonospora*. The DNA  
 CC encoding the biosynthetic proteins is useful for synthesising novel  
 CC everninomicin-related compounds, arising from modifications of the DNA  
 CC sequence designed to change glycosyl and modified orsellinic acid groups  
 CC contained in everninomicin, for expressing functional or mutant  
 CC everninomicin biosynthetic enzyme for evaluation, diagnosis and  
 CC preferably biosynthesis of everninomicin or other secondary metabolic  
 CC products, improving the yield of everninomicins and to produce novel  
 CC everninomicins and also as a hybridisation probe to identify homologous  
 CC sequences. The encoded polypeptides are useful for combinatorial  
 CC biosynthesis to generate libraries of orthomycins, e.g. everninomicin  
 CC analogues/homologues and drug discovery. The DNA encoding the integrase  
 CC allows for increasing a given gene dosage. The integrative vector can be  
 CC used to permanently integrate copies of a heterologous gene of choice  
 CC into chromosomes of different hosts and to integrate genes which increase  
 CC the yield of known products or to generate novel products such as hybrid  
 CC antibiotics or other novel secondary metabolites. The vector can also be  
 CC used to integrate antibiotic resistance genes in order to carry out  
 CC bioconversions with compounds to which the strain is normally sensitive  
 CC and is thus useful in fermentation processes involving e.g. *Streptomyces*  
 CC antibioticus. (Updated on 11-SEP-2003 to standardise OS field)

XX

SQ Sequence 474 AA;

Query Match 60.9%; Score 39; DB 4; Length 474;  
 Best Local Similarity 50.0%; Pred. No. 1.6e+02;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
 :|: :||| :||  
 Db 424 AGLKDTVYVSPK 435

# RESULT 8

ABP99325

ID ABP99325 standard; protein; 489 AA.

XX

AC ABP99325;

XX

DT 23-OCT-2003 (revised)

DT 21-MAR-2003 (first entry)

XX

DE Orthosomycin biosynthetic polypeptide SEQ ID NO 237.

XX

KW Orthosomycin; biosynthesis; everninomicin; avilamycin; enzyme.

XX

OS *Micromonospora carbonacea*; *aurantiaca*.

XX

PN WO200279505-A2.

XX

PD 10-OCT-2002.

XX

PF 28-MAR-2002; 2002WO-CA000432.

XX

PR 28-MAR-2001; 2001US-0279095P.

PR 30-MAR-2001; 2001US-0279709P.

PR 20-APR-2001; 2001US-0285214P.

XX  
PA (ECOP-) ECOPIA BIOSCIENCES INC.  
XX  
PI Farnet CM, Zazopoulos E, Staffa A;  
XX  
DR WPI; 2003-058435/05.  
DR N-PSDB; ABZ66788.  
XX  
PT Identifying orthosomycin biosynthetic gene, gene fragment or gene  
PT cluster, by detecting presence of nucleic acid sequence corresponding to  
PT 17 of flambamycins protein families.  
XX  
PS Claim 2; Page 387-389; 511pp; English.  
XX  
CC The invention relates to identifying orthosomycin biosynthetic genes and  
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin  
CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism. (Updated on  
CC 23-OCT-2003 to standardise OS field)  
XX  
SQ Sequence 489 AA;

Query Match 60.9%; Score 39; DB 6; Length 489;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
:|: :||| :||  
Db 439 TGLKDTVYVSPK 450

RESULT 9  
ABU31218  
ID ABU31218 standard; protein; 490 AA.  
XX  
AC ABU31218;  
XX  
DT 19-JUN-2003 (first entry)  
XX  
DE Protein encoded by Prokaryotic essential gene #16745.  
XX  
KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
XX  
OS Klebsiella pneumoniae.  
XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.

XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA35088.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 59142; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 490 AA;



Query Match 60.9%; Score 39; DB 6; Length 490;  
Best Local Similarity 66.7%; Pred. No. 1.7e+02;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
||| ||| |:  
Db 276 SGISNTSYSGSK 287

RESULT 10

ABP99326

ID ABP99326 standard; protein; 493 AA.

XX

AC ABP99326;

XX

DT 23-OCT-2003 (revised)

DT 21-MAR-2003 (first entry)

XX

DE Orthosomycin biosynthetic polypeptide SEQ ID NO 239.

XX

KW Orthosomycin; biosynthesis; everninomicin; avilamycin; enzyme.

XX

OS Micromonospora carbonacea; africana.

XX

PN WO200279505-A2.

XX

PD 10-OCT-2002.

XX

PF 28-MAR-2002; 2002WO-CA000432.

XX

PR 28-MAR-2001; 2001US-0279095P.

PR 30-MAR-2001; 2001US-0279709P.

PR 20-APR-2001; 2001US-0285214P.

XX

PA (ECOP-) ECOPIA BIOSCIENCES INC.

XX

PI Farnet CM, Zazopoulos E, Staffa A;

XX

DR WPI; 2003-058435/05.

DR N-PSDB; ABZ66789.

XX

PT Identifying orthosomycin biosynthetic gene, gene fragment or gene cluster, by detecting presence of nucleic acid sequence corresponding to 17 of flambamycins protein families.

XX

PS Claim 2; Page 390-391; 511pp; English.

XX

CC The invention relates to identifying orthosomycin biosynthetic genes and  
CC its fragment/gene cluster (ABZ66670-ABZ666813), comprising detecting the  
CC presence of a nucleic acid sequence coding for a polypeptide (ABP99207-  
CC ABP99362). The method is useful for identifying an orthosomycin  
CC biosynthetic gene, gene fragment or gene cluster, especially an  
CC everninomicin-type or avilamycin-type orthosomycin biosynthetic gene,  
CC gene fragment or gene cluster. The method is useful for detecting the  
CC presence of any organism that contains DNA for the production of  
CC orthosomycins (both everninomicin-type orthosomycins and avilamycin-type  
CC orthosomycins) regardless of the level at which genes for orthosomycin

CC production are expressed by the organism or the amount of orthosomycin  
CC produced by the organism. This allows for the detection of new  
CC orthosomycin natural products, not produced by the organism. (Updated on  
CC 23-OCT-2003 to standardise OS field)

XX

SQ Sequence 493 AA;

Query Match 60.9%; Score 39; DB 6; Length 493;  
Best Local Similarity 50.0%; Pred. No. 1.7e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
:|: :||| :||  
Db 443 AGLKDTVYVSPK 454

#### RESULT 11

ABO62726

ID ABO62726 standard; protein; 637 AA.

XX

AC ABO62726;

XX

DT 29-JUL-2004 (first entry)

XX

DE Klebsiella pneumoniae polypeptide seqid 9243.

XX

KW Recombinant expression vector; transcription regulatory element;

KW Klebsiella pneumoniae protein; antibacterial; Vaccine.

XX

OS Klebsiella pneumoniae.

XX

PN US6610836-B1.

XX

PD 26-AUG-2003.

XX

PF 27-JAN-2000; 2000US-00489039.

XX

PR 29-JAN-1999; 99US-0117747P.

XX

PA (GENC-) GENOME THERAPEUTICS CORP.

XX

PI Breton GL, Osborne M;

XX

DR WPI; 2003-895346/82.

DR N-PSDB; ACH96277.

XX

PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
PT preparing a vaccine composition against Klebsiella pneumoniae.

XX

PS Disclosure; SEQ ID NO 9243; 932pp; English.

XX

CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
CC pneumoniae polypeptide. Also described are: a recombinant expression  
CC vector comprising the nucleic acid, operably linked to a transcription  
CC regulatory element; and a cell comprising the recombinant expression  
CC vector. The nucleic acid is useful for preparing a vaccine composition  
CC against Klebsiella pneumoniae. This is the amino acid sequence of a

CC Klebsiella pneumoniae polypeptide of the invention

XX

SQ Sequence 637 AA;

Query Match 60.9%; Score 39; DB 7; Length 637;

Best Local Similarity 66.7%; Pred. No. 2.3e+02;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14

||||| |:

Db 423 SGISNTSYSGSK 434

#### RESULT 12

AAR30477

ID AAR30477 standard; protein; 858 AA.

XX

AC AAR30477;

XX

DT 25-MAR-2003 (revised)

DT 11-MAY-1993 (first entry)

XX

DE Human leukocyteHGF.

XX

KW Human; hepatocyte growth factor; recombinant; polylinker.

XX

OS Homo sapiens.

XX

FN WO9222321-A1.

XX

PD 23-DEC-1992.

XX

PF 19-MAY-1992; 92WO-US004227.

XX

PR 10-JUN-1991; 91US-00712284.

XX

PA (GETH ) GENENTECH INC.

XX

PI Jardieu PM;

XX

DR WPI; 1993-017907/02.

XX

PT Stimulating hepatocyte growth - comprises administering synergistic

PT amounts of hepatocyte growth factor and gamma-interferon, for treating

PT liver disease.

XX

PS Example 1; Page 19; 46pp; English.

XX

CC The plasmid pSVI6B5, which is a broadly applicable parental vector for  
CC expression of different polypeptides was derived from plasmid pSVI6B-tPA.  
CC pSVI6B5 (transformed E. coli strain ATCC No. 68, 151) carries polylinker  
CC regions in place of the t-PA cDNA in pSVI6B-tPA. These polylinker regions  
CC provide convenient, unique restriction endonuclease recognition sites  
CC that can be used to introduce any sequence that encodes a polypeptide of  
CC interest. Such a polypeptide is the human hepatocyte growth factor  
CC (hHGF). The DNA encoding hHGF may be isolated from a human leukocyte  
CC library and cloned into pSVI6B5 for expression of hHGF. (Updated on 25-

CC MAR-2003 to correct PN field.)  
XX  
SQ Sequence 858 AA;

Query Match 60.9%; Score 39; DB 2; Length 858;  
Best Local Similarity 63.6%; Pred. No. 3.2e+02;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 4 GISNTVYANPK 14  
|| | | ||  
Db 786 GIXNVTYNNPK 796

RESULT 13

AAW19604

ID AAW19604 standard; protein; 1024 AA.

XX

AC AAW19604;

XX

DT 21-AUG-1997 (first entry)

XX

DE Mycoplasma genitalium 116 kDa protein MG075 useful in vaccine.

XX

KW Mycoplasma; immunogen; vaccine; diagnosis; pneumonia; inflammation.

XX

OS Mycoplasma genitalium.

XX

PN WO9721727-A1.

XX

PD 19-JUN-1997.

XX

PF 13-DEC-1996; 96WO-AU000803.

XX

PR 13-DEC-1995; 95AU-00007127.

XX

PA (UYME ) UNIV MELBOURNE.

XX

PI Browning GF, Duffy MF, Whithear KG, Walker ID;

XX

DR WPI; 1997-332722/30.

XX

PT New immunogenic polypeptide(s) from Mycoplasma species - useful in  
PT vaccines and for diagnosis of Mycoplasma infection.

XX

PS Claim 19; Page 85-89; 110pp; English.

XX

CC Isolated or recombinant immunogenic polypeptides from Mycoplasma  
CC genitalium have mol.wt. of 16 kDa (AAW19603) (MG074) and 116 kDa  
CC (AAW19604) (MG075). They are homologues of 16 and 116 kDa proteins (see  
CC also AAW19601-02) obtd. from Mycoplasma pneumoniae. A genomic DNA  
CC sequence of M. genitalium contains contiguous open reading frames that  
CC code for the 2 polypeptides. Mycoplasma 16 or 116 kDa proteins, or  
CC immunogenic fragments that include a T or B cell epitope, can be used in  
CC vaccines for prevention and treatment of Mycoplasma infections, partic.  
CC in humans. They can also be used diagnostically to detect Mycoplasma, or  
CC to raise antibodies useful in immunoassays

XX

SQ Sequence 1024 AA;

Query Match 60.9%; Score 39; DB 2; Length 1024;  
Best Local Similarity 60.0%; Pred. No. 3.9e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ISNTVYANPK 14  
: |||::|||  
Db 773 LQNTVFSNPK 782

RESULT 14

AAG65232

ID AAG65232 standard; protein; 120 AA.

XX

AC AAG65232;

XX

DT 20-NOV-2001 (first entry)

XX

DE Human RNA helicase 13.

XX

KW Human; RNA helicase 13; cancer; haemopathy; HIV infection;  
KW immunological disease; inflammation; gene therapy.

XX

OS Homo sapiens.

XX

PN WO200165586-A1.

XX

PD 13-SEP-2001.

XX

PF 26-FEB-2001; 2001WO-CN000202.

XX

PR 07-MAR-2000; 2000CN-00111900.

XX

PA (BIO- ) BIOWINDOW GENE DEV INC SHANGHAI.

XX

PI Mao Y, Xie Y;

XX

DR WPI; 2001-565572/63.

DR N-PSDB; AAH79201.

XX

PT New human RNA helicase 13 for diagnosing and treating malignant neoplasm,  
PT hemopathy, human immunodeficiency virus infection, immunological diseases  
PT and various inflammations.

XX

PS Claim 1; Page 31; 37pp; Chinese.

XX

CC The present invention provides the protein and coding sequences of human  
CC RNA helicase 13. The sequences can be used in the treatment of cancer,  
CC haemopathy, HIV infection, immunological diseases and inflammation. The  
CC present sequence is the protein of the invention

XX

SQ Sequence 120 AA;

Query Match 59.4%; Score 38; DB 4; Length 120;  
Best Local Similarity 66.7%; Pred. No. 51;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
||||: | |||  
Db 12 SGISSVVSQNPK 23

RESULT 15

ABO00477

ID ABO00477 standard; protein; 329 AA.

XX

AC ABO00477;

XX

DT 06-AUG-2003 (first entry)

XX

DE Novel human polypeptide #64.

XX

KW Human; angiogenesis; cytokine; cell proliferation; pluripotent;  
KW cell differentiation; totipotent; stem cell; transplantation; bio-sensor;  
KW neuroepithelial cell; autoimmune disease; neural cell; genetic disorder;  
KW nerve; brain tissue; central nervous system disease;  
KW peripheral nervous system disease; neuropathy; haematopoiesis; bone;  
KW myeloid disorder; lymphoid cell disorder; platelet disorder; tendon;  
KW regeneration; cartilage; tendon; ligament; nerve tissue growth;  
KW tissue repair; wound healing; burn; ulcer; osteoporosis; cancer;  
KW osteoarthritis; bone degenerative disorder; periodontal disease;  
KW gut protection; lung fibrosis; liver fibrosis; reperfusion injury;  
KW immune deficiency; infection; autoimmune disorder; allergic reaction;  
KW thrombolysis; thrombosis; coagulation disorder; hereditary disorder;  
KW biorhythm; circadian cycle; fertility; metabolism; catabolism; anabolism;  
KW nootropic; neuroprotective; antiparkinsonian; anticonvulsant;  
KW haemostatic; vulnerary; antiulcer; osteopathic; antiarthritic;  
KW vasotropic; immunostimulant; antibacterial; fungicide; immunosuppressive;  
KW antirheumatic; antidiabetic; antiasthmatic; cytostatic; virucide.

XX

OS Homo sapiens.

XX

PN WO2003023013-A2.

XX

PD 20-MAR-2003.

XX

PF 13-SEP-2002; 2002WO-US029001.

XX

PR 13-SEP-2001; 2001US-0322511P.

PR 12-SEP-2002; 2002US-00243552.

XX

PA (HYSE-) HYSEQ INC.

XX

PI Tang YT, Yang Y, Wang Z, Weng G, Ma Y;

XX

DR WPI; 2003-313249/30.

DR N-PSDB; ACD05554.

XX

PT Novel nucleic acids and polypeptides for diagnosis, treatment of central  
PT and peripheral nervous system diseases and neuropathies, such as  
PT Alzheimer's, Parkinson's disease, Huntington's disease, amyotrophic  
PT lateral sclerosis.

XX

PS Claim 20; SEQ ID NO 400; 300pp; English.

XX

CC The present invention relates to the isolation of novel human  
CC polynucleotide sequences and their encoding polypeptides. The novel  
CC polypeptides exhibit activities relating to angiogenesis, cytokine, cell  
CC proliferation, cell differentiation, antiinflammatory, and stem cell  
CC growth factor activities. The polypeptides are involved in the  
CC proliferation, differentiation and survival of pluripotent and totipotent  
CC stem cells, and are useful for re-engineering damaged or diseased  
CC tissues, transplantation, manufacture of bio-pharmaceuticals and  
CC development of bio-sensors. The polypeptides can be used to manipulate  
CC stem cells in culture to give rise to neuroepithelial cells that can be  
CC used to augment or replace cells damaged by illness, autoimmune disease,  
CC accidental damage or genetic disorders. The polypeptides induce the  
CC proliferation of neural cells and regeneration of nerve and brain tissue  
CC and are useful for the treatment of central and peripheral nervous system  
CC diseases and neuropathies, such as Alzheimer's, Parkinson's disease,  
CC Huntington's disease, amyotrophic lateral sclerosis (ALS). The  
CC polypeptides are also involved in chemotactic or chemokinetic activity,  
CC regulation of haematopoiesis and are useful for treating myeloid or  
CC lymphoid cell disorders, platelet disorders such as thrombocytopaenia and  
CC for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue  
CC growth, in tissue repair, healing of burns, incisions, ulcers, for  
CC treating osteoporosis, osteoarthritis, bone degenerative disorders, and  
CC periodontal disease. The polypeptides are also useful for gut protection  
CC or regeneration and treatment of lung or liver fibrosis, reperfusion  
CC injury in various tissues, various immune deficiencies and disorders  
CC including severe combined immunodeficiency (SCID), bacterial or fungal  
CC infections, autoimmune disorders (e.g. multiple sclerosis, rheumatoid  
CC arthritis, diabetes mellitus, myasthenia gravis), allergic reactions and  
CC conditions, such as asthma or other respiratory problems. The  
CC polypeptides are involved in thrombolysis or thrombosis and are useful in  
CC treatment of various coagulation disorders (including hereditary  
CC disorders such as haemophilia) or to enhance coagulation and other  
CC haemostatic events in treating wounds resulting from trauma, surgery or  
CC other causes. The polypeptides exhibit immune stimulating or immune  
CC suppressing activity, and are useful for treating autoimmune diseases or  
CC cancer. They also inhibit the growth, infection or function of infectious  
CC agents such as bacteria, fungi, viruses, effect biorhythms or circadian  
CC cycles of rhythms, fertility of male or female subjects, metabolism,  
CC catabolism, and anabolism. ABO00414-ABO00749 represent the novel  
CC polypeptides of the invention. Note: The sequence data for this patent  
CC did not form part of the printed specification, but was obtained in  
CC electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 329 AA;

Query Match 59.4%; Score 38; DB 6; Length 329;  
Best Local Similarity 60.0%; Pred. No. 1.6e+02;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 ISNTVYANPK 14  
: | ||: ||  
Db 148 VENKVYSNPK 157

Search completed: January 31, 2005, 13:17:00  
Job time : 95.9545 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

CM protein - protein search, using sw model

Run on: January 31, 2005, 13:08:40 ; Search time 24.8182 Seconds  
(without alignments)  
37.410 Million cell updates/sec

Title: US-10-067-620-1  
Perfect score: 64  
Sequence: 1 XXSGISNTVYANPK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | %<br>Query Match | Length | DB | ID                   | Description       |
|------------|-------|------------------|--------|----|----------------------|-------------------|
| 1          | 39    | 60.9             | 106    | 4  | US-09-513-999C-8043  | Sequence 8043, Ap |
| 2          | 39    | 60.9             | 329    | 4  | US-09-107-532A-3780  | Sequence 3780, Ap |
| 3          | 39    | 60.9             | 637    | 4  | US-09-489-039A-9243  | Sequence 9243, Ap |
| 4          | 39    | 60.9             | 858    | 1  | US-07-712-284-2      | Sequence 2, Appli |
| 5          | 39    | 60.9             | 858    | 5  | PCT-US92-04227-2     | Sequence 2, Appli |
| 6          | 39    | 60.9             | 1024   | 3  | US-09-091-117-5      | Sequence 5, Appli |
| 7          | 38    | 59.4             | 66     | 4  | US-09-248-796A-21394 | Sequence 21394, A |
| 8          | 38    | 59.4             | 432    | 3  | US-09-118-319-2      | Sequence 2, Appli |
| 9          | 37    | 57.8             | 326    | 3  | US-09-286-691-23     | Sequence 23, Appl |



|    |    |      |      |   |                      |                   |
|----|----|------|------|---|----------------------|-------------------|
| 10 | 37 | 57.8 | 326  | 3 | US-09-687-147-23     | Sequence 23, Appl |
| 11 | 37 | 57.8 | 332  | 4 | US-09-428-034-4      | Sequence 4, Appli |
| 12 | 37 | 57.8 | 376  | 4 | US-09-428-034-2      | Sequence 2, Appli |
| 13 | 37 | 57.8 | 428  | 3 | US-09-118-319-5      | Sequence 5, Appli |
| 14 | 37 | 57.8 | 515  | 4 | US-09-107-532A-5317  | Sequence 5317, Ap |
| 15 | 37 | 57.8 | 700  | 4 | US-09-620-412C-345   | Sequence 345, App |
| 16 | 37 | 57.8 | 700  | 4 | US-09-598-419-345    | Sequence 345, App |
| 17 | 37 | 57.8 | 712  | 2 | US-08-468-576B-17    | Sequence 17, Appl |
| 18 | 37 | 57.8 | 712  | 2 | US-08-468-579B-17    | Sequence 17, Appl |
| 19 | 37 | 57.8 | 712  | 3 | US-08-468-577B-17    | Sequence 17, Appl |
| 20 | 37 | 57.8 | 841  | 4 | US-09-252-991A-33134 | Sequence 33134, A |
| 21 | 37 | 57.8 | 1752 | 4 | US-09-556-877-180    | Sequence 180, App |
| 22 | 37 | 57.8 | 1752 | 4 | US-09-620-412C-180   | Sequence 180, App |
| 23 | 37 | 57.8 | 1752 | 4 | US-09-598-419-180    | Sequence 180, App |
| 24 | 36 | 56.2 | 297  | 4 | US-09-107-532A-5306  | Sequence 5306, Ap |
| 25 | 36 | 56.2 | 525  | 4 | US-09-270-767-44715  | Sequence 44715, A |
| 26 | 36 | 56.2 | 1180 | 3 | US-09-224-024-28     | Sequence 28, Appl |
| 27 | 36 | 56.2 | 1180 | 5 | PCT-US94-07902-28    | Sequence 28, Appl |
| 28 | 35 | 54.7 | 159  | 4 | US-09-134-000C-4250  | Sequence 4250, Ap |
| 29 | 35 | 54.7 | 247  | 4 | US-09-248-796A-17530 | Sequence 17530, A |
| 30 | 35 | 54.7 | 256  | 4 | US-09-270-767-32493  | Sequence 32493, A |
| 31 | 35 | 54.7 | 256  | 4 | US-09-270-767-47710  | Sequence 47710, A |
| 32 | 35 | 54.7 | 284  | 4 | US-09-540-236-3124   | Sequence 3124, Ap |
| 33 | 35 | 54.7 | 442  | 3 | US-09-252-292C-29    | Sequence 29, Appl |
| 34 | 35 | 54.7 | 442  | 4 | US-09-567-615B-8     | Sequence 8, Appli |
| 35 | 35 | 54.7 | 455  | 4 | US-09-543-681A-8288  | Sequence 8288, Ap |
| 36 | 35 | 54.7 | 870  | 4 | US-09-543-681A-4768  | Sequence 4768, Ap |
| 37 | 35 | 54.7 | 1010 | 4 | US-09-248-796A-16379 | Sequence 16379, A |
| 38 | 35 | 54.7 | 1545 | 3 | US-08-296-791-4      | Sequence 4, Appli |
| 39 | 35 | 54.7 | 1545 | 4 | US-09-839-996-4      | Sequence 4, Appli |
| 40 | 35 | 54.7 | 1545 | 4 | US-10-080-505-4      | Sequence 4, Appli |
| 41 | 35 | 54.7 | 1545 | 5 | PCT-US95-10661A-4    | Sequence 4, Appli |
| 42 | 34 | 53.1 | 81   | 4 | US-09-248-796A-22138 | Sequence 22138, A |
| 43 | 34 | 53.1 | 98   | 4 | US-09-248-796A-23711 | Sequence 23711, A |
| 44 | 34 | 53.1 | 181  | 4 | US-09-270-767-57311  | Sequence 57311, A |
| 45 | 34 | 53.1 | 232  | 3 | US-09-134-001C-5367  | Sequence 5367, Ap |

#### ALIGNMENTS

##### RESULT 1

US-09-513-999C-8043

; Sequence 8043, Application US/09513999C

; Patent No. 6783961

; GENERAL INFORMATION:

; APPLICANT: Dumas Milne Edwards, J.B.

; APPLICANT: Duclert, A.

; APPLICANT: Giordano, J.Y.

; TITLE OF INVENTION: Expressed Sequence Tags and Encoded Human Proteins.

; Patent No. 6783961

; FILE REFERENCE: 59.US2.REG

; CURRENT APPLICATION NUMBER: US/09/513,999C

; CURRENT FILING DATE: 2000-02-24

; PRIOR APPLICATION NUMBER: US 60/122,487

; PRIOR FILING DATE: 1999-02-26

; NUMBER OF SEQ ID NOS: 36681

; SOFTWARE: Patent.pm  
; SEQ ID NO 8043  
; LENGTH: 106  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-513-999C-8043

Query Match 60.9%; Score 39; DB 4; Length 106;  
Best Local Similarity 63.6%; Pred. No. 8.7;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14  
|:| ||| ||:  
Db 23 GLSPTVYRNPE 33

RESULT 2

US-09-107-532A-3780

; Sequence 3780, Application US/09107532A  
; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND  
THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION  
; STREET: 100 Beaver Street  
; CITY: Waltham  
; STATE: Massachusetts  
; COUNTRY: USA  
; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660  
; COMPUTER: PC  
; OPERATING SYSTEM: <Unknown>  
; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107,532A  
; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598  
; FILING DATE: 14 May 1998  
; APPLICATION NUMBER: 60/051571  
; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariniello, Pamela Deneke  
; REGISTRATION NUMBER: 40,489  
; REFERENCE/DOCKET NUMBER: GTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007  
; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 3780:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 329 amino acids  
; TYPE: amino acid

```

;          TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          HYPOTHETICAL: YES
;          ORIGINAL SOURCE:
;          ORGANISM: Enterococcus faecium
;          FEATURE:
;          NAME/KEY: misc_feature
;          LOCATION: (B) LOCATION 1...329
;          SEQUENCE DESCRIPTION: SEQ ID NO: 3780:
US-09-107-532A-3780

```

```

Query Match          60.9%; Score 39; DB 4; Length 329;
Best Local Similarity 60.0%; Pred. No. 32;
Matches      6; Conservative 4; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy          3 SGISNTVYAN 12
             ||:||:|:|
Db          162 SGVSNSIHAN 171

```

#### RESULT 3

```

US-09-489-039A-9243
; Sequence 9243, Application US/09489039A
; Patent No. 6610836
; GENERAL INFORMATION:
; APPLICANT: Gary Breton et. al
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
KLEBSIELLA
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.2004001
; CURRENT APPLICATION NUMBER: US/09/489,039A
; CURRENT FILING DATE: 2000-01-27
; PRIOR APPLICATION NUMBER: US 60/117,747
; PRIOR FILING DATE: 1999-01-29
; NUMBER OF SEQ ID NOS: 14342
; SEQ ID NO 9243
; LENGTH: 637
; TYPE: PRT
; ORGANISM: Klebsiella pneumoniae
US-09-489-039A-9243

```

```

Query Match          60.9%; Score 39; DB 4; Length 637;
Best Local Similarity 66.7%; Pred. No. 67;
Matches      8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

```

```

Qy          3 SGISNTVYANPK 14
             ||||| |: |
Db          423 SGISNTSYSGSK 434

```

#### RESULT 4

```

US-07-712-284-2
; Sequence 2, Application US/07712284
; Patent No. 5227158
; GENERAL INFORMATION:
; APPLICANT: Jardieu, Paula M.
; TITLE OF INVENTION: Hepatocyte Growth Stimulation

```

```

; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
;   ADDRESSEE: Genentech, Inc.
;   STREET: 460 Point San Bruno Blvd
;   CITY: South San Francisco
;   STATE: California
;   COUNTRY: USA
;   ZIP: 94080
; COMPUTER READABLE FORM:
;   MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/07/712,284
;   FILING DATE: 19910610
;   CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
; ATTORNEY/AGENT INFORMATION:
;   NAME: Dreger, Ginger R.
;   REGISTRATION NUMBER: 33,055
;   REFERENCE/DOCKET NUMBER: 704
; TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 415/266-3216
;   TELEFAX: 415/952-9881
;   TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 858 amino acids
;     TYPE: AMINO ACID
;     TOPOLOGY: linear
US-07-712-284-2

```

```

Query Match          60.9%; Score 39; DB 1; Length 858;
Best Local Similarity 63.6%; Pred. No. 94;
Matches      7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      4 GISNTVYANPK 14
        || | | |||
Db      786 GIXNVTYNNPK 796

```

# RESULT 5

PCT-US92-04227-2

```

; Sequence 2, Application PC/TUS9204227
; GENERAL INFORMATION:
;   APPLICANT: GENENTECH, INC.
;   APPLICANT: Jardieu, Paula M.
;   TITLE OF INVENTION: Hepatocyte Growth Stimulation
;   NUMBER OF SEQUENCES: 2
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Genentech, Inc.
;     STREET: 460 Point San Bruno Blvd
;     CITY: South San Francisco
;     STATE: California

```

```

; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US92/04227
; FILING DATE: 19920519
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/712,284
; FILING DATE: 10 June 1991
; ATTORNEY/AGENT INFORMATION:
; NAME: Dreger, Ginger R.
; REGISTRATION NUMBER: 33,055
; REFERENCE/DOCKET NUMBER: 704P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-3216
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 858 amino acids
; TYPE: AMINO ACID
; TOPOLOGY: linear
PCT-US92-04227-2

```

```

Query Match          60.9%; Score 39; DB 5; Length 858;
Best Local Similarity 63.6%; Pred. No. 94;
Matches      7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

```

```

Qy      4 GISNTVYANPK 14
      || | | |||
Db      786 GIXNVTYNNPK 796

```

# RESULT 6

US-09-091-117-5

; Sequence 5, Application US/09091117

; Patent No. 6171589

; GENERAL INFORMATION:

; APPLICANT: The University of Melbourne

; TITLE OF INVENTION: Mycoplasma Recombinant Polypeptides and

; TITLE OF INVENTION: Vaccines

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GREENLEE, WINNER and SULLIVAN P.C.

; STREET: 5370 Manhattan Circle, Suite 201

; CITY: Boulder

; STATE: Colorado

; COUNTRY: United States of America

; ZIP: 80303

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

```

;   OPERATING SYSTEM:  PC-DOS/MS-DOS
;   SOFTWARE:  PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER:  US/09/091,117
;     FILING DATE:  12 JUNE 1998
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  PCT/AU96/00803
;     FILING DATE:  13-DEC-1996
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:  AU PN7127
;     FILING DATE:  13-DEC-1995
;   ATTORNEY/AGENT INFORMATION:
;     NAME:  WINNER, Ellen P.
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE:  +1 303 499 8080
;     TELEFAX:  +1 303 499 8089
;   INFORMATION FOR SEQ ID NO:  5:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH:  1024 amino acids
;       TYPE:  amino acid
;       TOPOLOGY:  linear
;     MOLECULE TYPE:  protein
;     ORIGINAL SOURCE:
;       ORGANISM:  Mycoplasma genitalium
US-09-091-117-5

```

```

Query Match          60.9%;  Score 39;  DB 3;  Length 1024;
Best Local Similarity 60.0%;  Pred. No. 1.1e+02;
Matches      6;  Conservative      3;  Mismatches      1;  Indels      0;  Gaps      0;

```

```

Qy      5 ISNTVYANPK 14
       : |||::|||
Db     773 LQNTVFSNPK 782

```

# RESULT 7

```

US-09-248-796A-21394
; Sequence 21394, Application US/09248796A
; Patent No. 6747137
; GENERAL INFORMATION:
;   APPLICANT: Keith Weinstock et al
;   TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO CANDIDA
ALBICANS
;   TITLE OF INVENTION:  FOR DIAGNOSTICS AND THERAPEUTICS
;   FILE REFERENCE: 107196.132
;   CURRENT APPLICATION NUMBER: US/09/248,796A
;   CURRENT FILING DATE:  1999-02-12
;   PRIOR APPLICATION NUMBER: US 60/074,725
;   PRIOR FILING DATE: 1998-02-13
;   PRIOR APPLICATION NUMBER: US 60/096,409
;   PRIOR FILING DATE: 1998-08-13
;   NUMBER OF SEQ ID NOS: 28208
;   SEQ ID NO 21394
;     LENGTH: 66
;     TYPE: PRT
;     ORGANISM: Candida albicans
US-09-248-796A-21394

```

Query Match 59.4%; Score 38; DB 4; Length 66;  
Best Local Similarity 80.0%; Pred. No. 7.7;  
Matches 8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYAN 12  
|| ||| |||  
Db 23 SGTSNTNYAN 32

RESULT 8

US-09-118-319-2

; Sequence 2, Application US/09118319  
; Patent No. 6114158  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xin-Liang  
; APPLICANT: Chen, Huizhong  
; APPLICANT: Ljungdahl, Lars G.  
; TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences  
; FILE REFERENCE: 33-98sequence listing  
; CURRENT APPLICATION NUMBER: US/09/118,319  
; CURRENT FILING DATE: 1998-07-17  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 2  
; LENGTH: 432  
; TYPE: PR1  
; ORGANISM: Orpinomyces sp. PC-2  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence:oligonucleotide  
US-09-118-319-2

Query Match 59.4%; Score 38; DB 3; Length 432;  
Best Local Similarity 66.7%; Pred. No. 65;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 6 SNTVYANPK 14  
:| :|||  
Db 109 TNQIYANPK 117

RESULT 9

US-09-286-691-23

; Sequence 23, Application US/09286691  
; Patent No. 6190189  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xin-Liang  
; APPLICANT: Ljungdahl, Lars G.  
; APPLICANT: Chen, Huizhong  
; TITLE OF INVENTION: Cellulases and Coding Sequences  
; FILE REFERENCE: 42-96  
; CURRENT APPLICATION NUMBER: US/09/286,691  
; CURRENT FILING DATE: 1999-04-05  
; EARLIER APPLICATION NUMBER: US 60/027,883  
; EARLIER FILING DATE: 1996-10-04  
; EARLIER APPLICATION NUMBER: PCT US97/18008  
; EARLIER FILING DATE: 1997-10-03

; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 326  
; TYPE: PRT  
; ORGANISM: Neocallimastix patriciarum  
US-09-286-691-23

Query Match 57.8%; Score 37; DB 3; Length 326;  
Best Local Similarity 75.0%; Pred. No. 72;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 7 NTVYANPK 14  
| :|||||  
Db 9 NQIYANPK 16

RESULT 10

US-09-687-147-23  
; Sequence 23, Application US/09687147  
; Patent No. 6268198  
; GENERAL INFORMATION:  
; APPLICANT: Li, Xin-Liang  
; APPLICANT: Ljungdahl, Lars G.  
; APPLICANT: Chen, Huizhong  
; TITLE OF INVENTION: Cellulases and Coding Sequences  
; FILE REFERENCE: 42-96a  
; CURRENT APPLICATION NUMBER: US/09/687,147  
; CURRENT FILING DATE: 2000-10-12  
; PRIOR APPLICATION NUMBER: US 60/027,883  
; PRIOR FILING DATE: 1996-10-04  
; PRIOR APPLICATION NUMBER: PCT US97/18008  
; PRIOR FILING DATE: 1997-10-03  
; PRIOR APPLICATION NUMBER: 09/286,691  
; PRIOR FILING DATE: 1999-04-05  
; NUMBER OF SEQ ID NOS: 29  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 23  
; LENGTH: 326  
; TYPE: PRT  
; ORGANISM: Neocallimastix patriciarum  
US-09-687-147-23

Query Match 57.8%; Score 37; DB 3; Length 326;  
Best Local Similarity 75.0%; Pred. No. 72;  
Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 7 NTVYANPK 14  
| :|||||  
Db 9 NQIYANPK 16

RESULT 11

US-09-428-034-4  
; Sequence 4, Application US/09428034  
; Patent No. 6428996  
; GENERAL INFORMATION:



; APPLICANT: Cheng, Kuo-Joan  
 ; APPLICANT: Liu, Jin-Hao  
 ; APPLICANT: Tsai, Cheng-Fang  
 ; APPLICANT: Hsu, Yih-Chin  
 ; TITLE OF INVENTION: CELLULASE ENZYMES  
 ; FILE REFERENCE: 08919/036001  
 ; CURRENT APPLICATION NUMBER: US/09/428,034  
 ; CURRENT FILING DATE: 1999-10-27  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 4  
 ; LENGTH: 332  
 ; TYPE: PRT  
 ; ORGANISM: *Piromyces rhizinflata*  
 US-09-428-034-4

Query Match 57.8%; Score 37; DB 4; Length 332;  
 Best Local Similarity 75.0%; Pred. No. 74;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 7 NTVYANPK 14  
 | :|||||  
 Db 11 NEIYANPK 18

# RESULT 12

US-09-428-034-2  
 ; Sequence 2, Application US/09428034  
 ; Patent No. 6428996  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Cheng, Kuo-Joan  
 ; APPLICANT: Liu, Jin-Hao  
 ; APPLICANT: Tsai, Cheng-Fang  
 ; APPLICANT: Hsu, Yih-Chin  
 ; TITLE OF INVENTION: CELLULASE ENZYMES  
 ; FILE REFERENCE: 08919/036001  
 ; CURRENT APPLICATION NUMBER: US/09/428,034  
 ; CURRENT FILING DATE: 1999-10-27  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: FastSEQ for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 376  
 ; TYPE: PRT  
 ; ORGANISM: *Piromyces rhizinflata*  
 US-09-428-034-2

Query Match 57.8%; Score 37; DB 4; Length 376;  
 Best Local Similarity 75.0%; Pred. No. 85;  
 Matches 6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 7 NTVYANPK 14  
 | :|||||  
 Db 54 NEIYANPK 61

# RESULT 13

US-09-118-319-5

```
; Sequence 5, Application US/09118319
; Patent No. 6114158
; GENERAL INFORMATION:
; APPLICANT: Li, Xin-Liang
; APPLICANT: Chen, Huizhong
; APPLICANT: Ljungdahl, Lars G.
; TITLE OF INVENTION: Orpinomyces Cellulase Celf Protein and Coding Sequences
; FILE REFERENCE: 33-98sequence listing
; CURRENT APPLICATION NUMBER: US/09/118,319
; CURRENT FILING DATE: 1998-07-17
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 5
;   LENGTH: 428
;   TYPE: PR1
;   ORGANISM: Neocallimastix patriciarum
US-09-118-319-5
```

```
Query Match          57.8%; Score 37; DB 3; Length 428;
Best Local Similarity 75.0%; Pred. No. 99;
Matches      6; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy      7 NTVYANPK 14
        | :|||||
Db     111 NQIYANPK 118
```

#### RESULT 14

US-09-107-532A-5317

```
; Sequence 5317, Application US/09107532A
; Patent No. 6583275
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND
THERAPEUTICS
; NUMBER OF SEQUENCES: 7310
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
```

```

; ATTORNEY/AGENT INFORMATION:
;     NAME: Ariniello, Pamela Deneke
;     REGISTRATION NUMBER: 40,489
;     REFERENCE/DOCKET NUMBER: GTC-012
; TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (781)893-5007
;     TELEFAX: (781)893-8277
; INFORMATION FOR SEQ ID NO: 5317:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 515 amino acids
;         TYPE: amino acid
;         TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     HYPOTHETICAL: YES
;     ORIGINAL SOURCE:
;         ORGANISM: Enterococcus faecium
;     FEATURE:
;         NAME/KEY: misc_feature
;         LOCATION: (B) LOCATION 1...515
;     SEQUENCE DESCRIPTION: SEQ ID NO: 5317:
US-09-107-532A-5317

```

```

Query Match          57.8%;  Score 37;  DB 4;  Length 515;
Best Local Similarity 60.0%;  Pred. No. 1.2e+02;
Matches      6;  Conservative  4;  Mismatches    0;  Indels    0;  Gaps    0;

```

```

Qy      5 ISNTVYANPK 14
        |::|::|
Db      440 IADTLFANPK 449

```

```

RESULT 15
US-09-620-412C-345
; Sequence 345, Application US/09620412C
; Patent No. 6448234
; GENERAL INFORMATION:
;   APPLICANT: Steven P. Fling
;   TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR TREATMENT AND
;   TITLE OF INVENTION: DIAGNOSIS OF CHLAMYDIAL INFECTION
;   FILE REFERENCE: 210121.469C7
;   CURRENT APPLICATION NUMBER: US/09/620,412C
;   CURRENT FILING DATE: 2000-07-20
;   NUMBER OF SEQ ID NOS: 363
;   SOFTWARE: FastSEQ for Windows Version 3.0/4.0
;   SEQ ID NO 345
;     LENGTH: 700
;     TYPE: PRT
;     ORGANISM: Chlamydia trachomatis
US-09-620-412C-345

```

```

Query Match          57.8%;  Score 37;  DB 4;  Length 700;
Best Local Similarity 50.0%;  Pred. No. 1.7e+02;
Matches      6;  Conservative  4;  Mismatches    2;  Indels    0;  Gaps    0;

```

```

Qy      3 SGISNTVYANPK 14
        ||:|::: ||
Db      252 SGVSSSIPTNPK 263

```

Search completed: January 31, 2005, 13:25:08  
Job time : 25.8182 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 13:22:56 ; Search time 79.8636 Seconds  
(without alignments)  
63.334 Million cell updates/sec

Title: US-10-067-620-1  
Perfect score: 64  
Sequence: 1 XXSGISNTVYANPK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*
- 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*
- 17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*
- 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*
- 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*
- 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,

and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result<br>No. | Score | %     |              | DB | ID                   | Description       |
|---------------|-------|-------|--------------|----|----------------------|-------------------|
|               |       | Query | Match Length |    |                      |                   |
| 1             | 62    | 96.9  | 14           | 14 | US-10-067-484-1      | Sequence 1, Appli |
| 2             | 62    | 96.9  | 14           | 14 | US-10-067-620-1      | Sequence 1, Appli |
| 3             | 40    | 62.5  | 266          | 15 | US-10-424-599-221528 | Sequence 221528,  |
| 4             | 39    | 60.9  | 401          | 10 | US-09-769-734-7      | Sequence 7, Appli |
| 5             | 39    | 60.9  | 474          | 11 | US-09-758-759-39     | Sequence 39, Appl |
| 6             | 39    | 60.9  | 489          | 14 | US-10-107-431-237    | Sequence 237, App |
| 7             | 39    | 60.9  | 490          | 15 | US-10-282-122A-59142 | Sequence 59142, A |
| 8             | 39    | 60.9  | 493          | 14 | US-10-107-431-239    | Sequence 239, App |
| 9             | 39    | 60.9  | 1522         | 14 | US-10-369-493-3471   | Sequence 3471, Ap |
| 10            | 38    | 59.4  | 329          | 14 | US-10-243-552-494    | Sequence 494, App |
| 11            | 38    | 59.4  | 336          | 15 | US-10-282-122A-77223 | Sequence 77223, A |
| 12            | 38    | 59.4  | 552          | 14 | US-10-369-493-20174  | Sequence 20174, A |
| 13            | 38    | 59.4  | 552          | 16 | US-10-322-281-732    | Sequence 732, App |
| 14            | 38    | 59.4  | 697          | 16 | US-10-322-281-729    | Sequence 729, App |
| 15            | 37    | 57.8  | 586          | 14 | US-10-128-714-3565   | Sequence 3565, Ap |
| 16            | 37    | 57.8  | 586          | 14 | US-10-128-714-8565   | Sequence 8565, Ap |
| 17            | 37    | 57.8  | 700          | 9  | US-09-841-132-345    | Sequence 345, App |
| 18            | 37    | 57.8  | 700          | 17 | US-10-872-155-345    | Sequence 345, App |
| 19            | 37    | 57.8  | 830          | 15 | US-10-282-122A-43637 | Sequence 43637, A |
| 20            | 37    | 57.8  | 831          | 15 | US-10-282-122A-69302 | Sequence 69302, A |
| 21            | 37    | 57.8  | 955          | 16 | US-10-437-963-149675 | Sequence 149675,  |
| 22            | 37    | 57.8  | 1430         | 16 | US-10-437-963-149674 | Sequence 149674,  |
| 23            | 37    | 57.8  | 1751         | 9  | US-09-841-132-445    | Sequence 445, App |
| 24            | 37    | 57.8  | 1751         | 9  | US-09-841-132-594    | Sequence 594, App |
| 25            | 37    | 57.8  | 1751         | 16 | US-10-467-534-45     | Sequence 45, Appl |
| 26            | 37    | 57.8  | 1751         | 17 | US-10-872-155-445    | Sequence 445, App |
| 27            | 37    | 57.8  | 1751         | 17 | US-10-872-155-594    | Sequence 594, App |
| 28            | 37    | 57.8  | 1752         | 9  | US-09-841-132-180    | Sequence 180, App |
| 29            | 37    | 57.8  | 1752         | 17 | US-10-872-155-180    | Sequence 180, App |
| 30            | 36    | 56.2  | 118          | 16 | US-10-437-963-177163 | Sequence 177163,  |
| 31            | 36    | 56.2  | 125          | 15 | US-10-424-599-272203 | Sequence 272203,  |
| 32            | 36    | 56.2  | 201          | 9  | US-09-864-761-37006  | Sequence 37006, A |
| 33            | 36    | 56.2  | 278          | 15 | US-10-424-599-255418 | Sequence 255418,  |
| 34            | 36    | 56.2  | 281          | 15 | US-10-425-114-63633  | Sequence 63633, A |
| 35            | 36    | 56.2  | 287          | 16 | US-10-408-765A-1376  | Sequence 1376, Ap |
| 36            | 36    | 56.2  | 303          | 9  | US-09-882-837-2      | Sequence 2, Appli |
| 37            | 36    | 56.2  | 303          | 14 | US-10-175-696-29     | Sequence 29, Appl |
| 38            | 36    | 56.2  | 303          | 14 | US-10-220-380-1      | Sequence 1, Appli |
| 39            | 36    | 56.2  | 303          | 16 | US-10-776-871-29     | Sequence 29, Appl |
| 40            | 36    | 56.2  | 370          | 15 | US-10-282-122A-48868 | Sequence 48868, A |
| 41            | 36    | 56.2  | 376          | 15 | US-10-425-114-46448  | Sequence 46448, A |
| 42            | 36    | 56.2  | 381          | 17 | US-10-425-115-188855 | Sequence 188855,  |
| 43            | 36    | 56.2  | 389          | 15 | US-10-425-114-65684  | Sequence 65684, A |
| 44            | 36    | 56.2  | 465          | 17 | US-10-425-115-252132 | Sequence 252132,  |
| 45            | 36    | 56.2  | 468          | 15 | US-10-282-122A-67947 | Sequence 67947, A |

#### ALIGNMENTS

RESULT 1

US-10-067-484-1

```
; Sequence 1, Application US/10067484
; Publication No. US20030170763A1
; GENERAL INFORMATION:
; APPLICANT: Buchanan, Bob B.
; APPLICANT: del Val, Gregorio
; APPLICANT: Frick, Oscar L.
; TITLE OF INVENTION: RAGWEED ALLERGENS
; FILE REFERENCE: 416272000200
; CURRENT APPLICATION NUMBER: US/10/067,484
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 60/266,686
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Ragweed
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: 1,2
; OTHER INFORMATION: Xaa = Leucine or Isoleucine
US-10-067-484-1
```

```
Query Match          96.9%; Score 62; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 0.00013;
Matches 12; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      3 SGISNTVYANPK 14
        |||||
Db      3 SGISNTVYANPK 14
```

RESULT 2

US-10-067-620-1

```
; Sequence 1, Application US/10067620
; Publication No. US20030180225A1
; GENERAL INFORMATION:
; APPLICANT: Buchanan, Bob B.
; APPLICANT: del Val, Gregorio
; APPLICANT: Frick, Oscar L.
; APPLICANT: Teuber, Suzanne S.
; TITLE OF INVENTION: WALNUT AND RYEGRASS ALLERGENS
; FILE REFERENCE: 416272003400
; CURRENT APPLICATION NUMBER: US/10/067,620
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 60/266,686
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Ragweed
; FEATURE:
```



; TITLE OF INVENTION: Genetic Locus for Everninomicin Biosynthesis  
; FILE REFERENCE: PA 005-US  
; CURRENT APPLICATION NUMBER: US/09/769,734  
; CURRENT FILING DATE: 2001-01-26  
; NUMBER OF SEQ ID NOS: 58  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 7  
; LENGTH: 401  
; TYPE: PRT  
; ORGANISM: M. carbonacea  
US-09-769-734-7

Query Match 60.9%; Score 39; DB 10; Length 401;  
Best Local Similarity 50.0%; Pred. No. 1.1e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
:|: :||| :||  
Db 351 TGLKDTVYVSPK 362

RESULT 5

US-09-758-759-39  
; Sequence 39, Application US/09758759  
; Publication No. US20040101832A1  
; GENERAL INFORMATION:  
; APPLICANT: Hosted, Thomas J.  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Everninomicin Biosynthetic Genes  
; FILE REFERENCE: ID0983K US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 39  
; LENGTH: 474  
; TYPE: PRT  
; ORGANISM: Micromonospora carbonacea  
; FEATURE:  
; OTHER INFORMATION: evrG  
US-09-758-759-39

Query Match 60.9%; Score 39; DB 11; Length 474;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
:|: :||| :||  
Db 424 AGLKDTVYVSPK 435

RESULT 6

US-10-107-431-237  
; Sequence 237, Application US/10107431



; Publication No. US20030224364A1  
; GENERAL INFORMATION:  
; APPLICANT: Farnet, Chris  
; APPLICANT: Staffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND  
DISTINGUISHING ORTHOSOMYCIN  
; TITLE OF INVENTION: BIOSYNTHETIC LOCI  
; FILE REFERENCE: 3001-7US  
; CURRENT APPLICATION NUMBER: US/10/107,431  
; CURRENT FILING DATE: 2002-03-28  
; NUMBER OF SEQ ID NOS: 282  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 237  
; LENGTH: 489  
; TYPE: PRT  
; ORGANISM: Micromonospora carbonacea aurantiaca  
US-10-107-431-237

Query Match 60.9%; Score 39; DB 14; Length 489;  
Best Local Similarity 50.0%; Pred. No. 1.3e+02;  
Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
:|: :||| :||  
Db 439 TGLKDTVYVSPK 450

# RESULT 7

US-10-282-122A-59142  
; Sequence 59142, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Malone, Cheryl  
; APPLICANT: Haselbeck, Robert  
; APPLICANT: Ohlsen, Kari  
; APPLICANT: Zyskind, Judith  
; APPLICANT: Wall, Daniel  
; APPLICANT: Trawick, John  
; APPLICANT: Carr, Grant  
; APPLICANT: Yamamoto, Robert  
; APPLICANT: Forsyth, R.  
; APPLICANT: Xu, H.  
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
; FILE REFERENCE: ELITRA.034A  
; CURRENT APPLICATION NUMBER: US/10/282,122A  
; CURRENT FILING DATE: 2003-02-20  
; PRIOR APPLICATION NUMBER: 60/191,078  
; PRIOR FILING DATE: 2000-03-21  
; PRIOR APPLICATION NUMBER: 60/206,848  
; PRIOR FILING DATE: 2000-05-23  
; PRIOR APPLICATION NUMBER: 60/207,727  
; PRIOR FILING DATE: 2000-05-26  
; PRIOR APPLICATION NUMBER: 60/230,335  
; PRIOR FILING DATE: 2000-09-06

```
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 59142
;   LENGTH: 490
;   TYPE: PRT
;   ORGANISM: Klebsiella pneumoniae
US-10-282-122A-59142
```

```
Query Match          60.9%; Score 39; DB 15; Length 490;
Best Local Similarity 66.7%; Pred. No. 1.3e+02;
Matches      8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;
```

```
Qy      3 SGISNTVYANPK 14
        ||||| |: |
Db      276 SGISNTSYSGSK 287
```

# RESULT 8

```
US-10-107-431-239
; Sequence 239, Application US/10107431
; Publication No. US20030224364A1
; GENERAL INFORMATION:
; APPLICANT: Farnet, Chris
; APPLICANT: Staffa, Alfredo
; APPLICANT: Zazopoulos, Emmanuel
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR IDENTIFYING AND
DISTINGUISHING ORTHOSOMYCIN
; TITLE OF INVENTION: BIOSYNTHETIC LOCI
; FILE REFERENCE: 3001-7US
; CURRENT APPLICATION NUMBER: US/10/107,431
; CURRENT FILING DATE: 2002-03-28
; NUMBER OF SEQ ID NOS: 282
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 239
;   LENGTH: 493
;   TYPE: PRT
;   ORGANISM: Micromonospora carbonacea africana
US-10-107-431-239
```

```
Query Match          60.9%; Score 39; DB 14; Length 493;
Best Local Similarity 50.0%; Pred. No. 1.3e+02;
Matches      6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      3 SGISNTVYANPK 14
```

Db                   :|: :||| :||  
443 AGLKDTVYVSPK 454

RESULT 9

US-10-369-493-3471

; Sequence 3471, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 3471

; LENGTH: 1522

; TYPE: PRT

; ORGANISM: Neurospora crassa

US-10-369-493-3471

Query Match                   60.9%;   Score 39;   DB 14;   Length 1522;  
Best Local Similarity   58.3%;   Pred. No. 4.7e+02;  
Matches       7;   Conservative   3;   Mismatches   2;   Indels    0;   Gaps    0;

Qy                   3 SGISNTVYANPK 14  
                     :|||||: | |:  
Db                   489 AGISNTISATPE 500

RESULT 10

US-10-243-552-494

; Sequence 494, Application US/10243552

; Publication No. US20030224379A1

; GENERAL INFORMATION:

; APPLICANT: Tang, Y. Tom

; APPLICANT: Yang, Yonghong

; APPLICANT: Wang, Zhiwei

; APPLICANT: Weng, Gezhi

; APPLICANT: Ma, Yunqing

; TITLE OF INVENTION: Novel Nucleic Acids and

; TITLE OF INVENTION: Polypeptides

; FILE REFERENCE: 807A

; CURRENT APPLICATION NUMBER: US/10/243,552

; CURRENT FILING DATE: 2002-09-12

; PRIOR APPLICATION NUMBER: US 60/322,511

; PRIOR FILING DATE: 2001-09-13

; PRIOR APPLICATION NUMBER: PCT/US00/35017

; PRIOR FILING DATE: 2000-12-22

; PRIOR APPLICATION NUMBER: US 09/488,725  
 ; PRIOR FILING DATE: 2000-01-21  
 ; PRIOR APPLICATION NUMBER: US 09/552,317  
 ; PRIOR FILING DATE: 2000-04-25  
 ; PRIOR APPLICATION NUMBER: PCT/US01/02623  
 ; PRIOR FILING DATE: 2001-01-25  
 ; PRIOR APPLICATION NUMBER: US 09/491,404  
 ; PRIOR FILING DATE: 2000-01-25  
 ; PRIOR APPLICATION NUMBER: PCT/US01/03800  
 ; PRIOR FILING DATE: 2001-02-05  
 ; PRIOR APPLICATION NUMBER: US 09/496,914  
 ; PRIOR FILING DATE: 2000-02-03  
 ; PRIOR APPLICATION NUMBER: US 09/560,875  
 ; PRIOR FILING DATE: 2000-04-27  
 ; PRIOR APPLICATION NUMBER: PCT/US01/04927  
 ; PRIOR FILING DATE: 2001-02-26  
 ; Remaining Prior Application data removed - See File Wrapper or PALM.  
 ; NUMBER OF SEQ ID NOS: 998  
 ; SOFTWARE: pt\_FL\_genes Version 5.0  
 ; SEQ ID NO 494  
 ; LENGTH: 329  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 US-10-243-552-494

Query Match 59.4%; Score 38; DB 14; Length 329;  
 Best Local Similarity 60.0%; Pred. No. 1.3e+02;  
 Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 5 ISNTVYANPK 14  
 : | ||: ||  
 Db 148 VENKVYSNPK 157

RESULT 11  
 US-10-282-122A-77223  
 ; Sequence 77223, Application US/10282122A  
 ; Publication No. US20040029129A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Wang, Liangsu  
 ; APPLICANT: Zamudio, Carlos  
 ; APPLICANT: Malone, Cheryl  
 ; APPLICANT: Haselbeck, Robert  
 ; APPLICANT: Ohlsen, Kari  
 ; APPLICANT: Zyskind, Judith  
 ; APPLICANT: Wall, Daniel  
 ; APPLICANT: Trawick, John  
 ; APPLICANT: Carr, Grant  
 ; APPLICANT: Yamamoto, Robert  
 ; APPLICANT: Forsyth, R.  
 ; APPLICANT: Xu, H.  
 ; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms  
 ; FILE REFERENCE: ELITRA.034A  
 ; CURRENT APPLICATION NUMBER: US/10/282,122A  
 ; CURRENT FILING DATE: 2003-02-20  
 ; PRIOR APPLICATION NUMBER: 60/191,078  
 ; PRIOR FILING DATE: 2000-03-21

```

; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 77223
;   LENGTH: 336
;   TYPE: PRT
;   ORGANISM: Vibrio cholerae
US-10-282-122A-77223

```

```

Query Match          59.4%; Score 38; DB 15; Length 336;
Best Local Similarity 54.5%; Pred. No. 1.3e+02;
Matches      6; Conservative    3; Mismatches    2; Indels      0; Gaps      0;

```

```

Qy      4 GISNTVYANPK 14
        |:|  ::|||
Db      250 GVSKELFANPK 260

```

# RESULT 12

```

US-10-369-493-20174
; Sequence 20174, Application US/10369493
; Publication No. US20030233675A1
; GENERAL INFORMATION:
; APPLICANT: Cao, Yongwei
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Slater, Steven C.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Chen, Xianfeng
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION
OF
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES
; FILE REFERENCE: 38-10(52052)B
; CURRENT APPLICATION NUMBER: US/10/369,493
; CURRENT FILING DATE: 2003-02-28
; PRIOR APPLICATION NUMBER: US 60/360,039
; PRIOR FILING DATE: 2002-02-21
; NUMBER OF SEQ ID NOS: 47374
; SEQ ID NO 20174
;   LENGTH: 552

```

; TYPE: PRT  
; ORGANISM: No. US20030233675Altoc punctiforme  
US-10-369-493-20174

Query Match 59.4%; Score 38; DB 14; Length 552;  
Best Local Similarity 45.5%; Pred. No. 2.3e+02;  
Matches 5; Conservative 5; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14  
|::|::|||:  
Db 530 GLTKTIFANPQ 540

RESULT 13

US-10-322-281-732

; Sequence 732, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 732  
; LENGTH: 552  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-322-281-732

Query Match 59.4%; Score 38; DB 16; Length 552;  
Best Local Similarity 60.0%; Pred. No. 2.3e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 SGISNTVYAN 12  
||::|:| |  
Db 252 SGVNNSVYTN 261

RESULT 14

US-10-322-281-729

; Sequence 729, Application US/10322281  
; Publication No. US20040126762A1  
; GENERAL INFORMATION:  
; APPLICANT: David W. Morris  
; APPLICANT: Marc S. Malandro  
; TITLE OF INVENTION: Novel Compositions and Methods in Cancer  
; FILE REFERENCE: 529452001000  
; CURRENT APPLICATION NUMBER: US/10/322,281  
; CURRENT FILING DATE: 2002-12-17  
; NUMBER OF SEQ ID NOS: 866  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 729  
; LENGTH: 697  
; TYPE: PRT

; ORGANISM: Mus musculus  
US-10-322-281-729

Query Match 59.4%; Score 38; DB 16; Length 697;  
Best Local Similarity 60.0%; Pred. No. 3e+02;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 3 SGISNTVYAN 12  
||::||:|  
Db 441 SGVNNSVYTN 450

RESULT 15

US-10-128-714-3565

; Sequence 3565, Application US/10128714  
; Publication No. US20030119013A1  
; GENERAL INFORMATION:  
; APPLICANT: Jiang, Bo  
; APPLICANT: Hu, Wenqi  
; APPLICANT: Tishkoff, Daniel  
; APPLICANT: Zamudio, Carlos  
; APPLICANT: Eroshkin, Alexey M  
; APPLICANT: Lemieux, Sebastien M  
; TITLE OF INVENTION: Identification of Essential Genes in Aspergillus  
fumigatus and  
; TITLE OF INVENTION: Methods of Use  
; FILE REFERENCE: 10182-018-999  
; CURRENT APPLICATION NUMBER: US/10/128,714  
; CURRENT FILING DATE: 2002-04-23  
; PRIOR APPLICATION NUMBER: US 60/285,697  
; PRIOR FILING DATE: 2001-04-23  
; PRIOR APPLICATION NUMBER: US 60/287,066  
; PRIOR FILING DATE: 2001-04-27  
; PRIOR APPLICATION NUMBER: US 60/295,890  
; PRIOR FILING DATE: 2001-06-05  
; PRIOR APPLICATION NUMBER: US 60/303,899  
; PRIOR FILING DATE: 2001-07-09  
; PRIOR APPLICATION NUMBER: US 60/316,362  
; PRIOR FILING DATE: 2001-08-31  
; NUMBER OF SEQ ID NOS: 8603  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 3565  
; LENGTH: 586  
; TYPE: PRT  
; ORGANISM: Aspergillus fumigatus  
US-10-128-714-3565

Query Match 57.8%; Score 37; DB 14; Length 586;  
Best Local Similarity 58.3%; Pred. No. 3.8e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
|::||||:|  
Db 495 SEMTNTVYDDPK 506

Search completed: January 31, 2005, 13:44:49

Job time : 80.8636 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 13:07:55 ; Search time 18.4545 Seconds  
(without alignments)  
72.992 Million cell updates/sec

Title: US-10-067-620-1  
Perfect score: 64  
Sequence: 1 XXSGISNTVYANPK 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result<br>No. | Score | %<br>Query |        | DB | ID     | Description        |
|---------------|-------|------------|--------|----|--------|--------------------|
|               |       | Match      | Length |    |        |                    |
| 1             | 41    | 64.1       | 382    | 2  | E83846 | D-alanyl-D-alanine |
| 2             | 39    | 60.9       | 673    | 2  | T47905 | hypothetical prote |
| 3             | 39    | 60.9       | 1024   | 2  | C64208 | hypothetical prote |
| 4             | 38    | 59.4       | 336    | 2  | F82242 | oligopeptide ABC t |
| 5             | 37    | 57.8       | 93     | 2  | D69262 | hypothetical prote |
| 6             | 37    | 57.8       | 339    | 2  | E84671 | hypothetical prote |
| 7             | 37    | 57.8       | 389    | 2  | H81146 | penicillin-binding |
| 8             | 37    | 57.8       | 389    | 2  | D81875 | probable serine-ty |
| 9             | 37    | 57.8       | 562    | 2  | S44287 | pyruvate kinase, p |
| 10            | 37    | 57.8       | 653    | 2  | F70383 | organic solvent to |
| 11            | 37    | 57.8       | 830    | 2  | E83031 | conserved hypothet |
| 12            | 37    | 57.8       | 1751   | 2  | G71518 | hypothetical prote |



|    |      |      |      |   |        |                    |
|----|------|------|------|---|--------|--------------------|
| 13 | 36   | 56.2 | 117  | 2 | C64378 | hypothetical prote |
| 14 | 36   | 56.2 | 205  | 2 | AC3366 | serine-type D-Ala- |
| 15 | 36   | 56.2 | 243  | 2 | S43887 | restriction endonu |
| 16 | 36   | 56.2 | 243  | 2 | F81130 | type II restrictio |
| 17 | 36   | 56.2 | 507  | 2 | E64247 | phosphoglycerate m |
| 18 | 36   | 56.2 | 969  | 2 | T15446 | hypothetical prote |
| 19 | 36   | 56.2 | 1180 | 2 | A26858 | parasporal crystal |
| 20 | 36   | 56.2 | 1180 | 2 | I39870 | parasporal crystal |
| 21 | 35.5 | 55.5 | 981  | 2 | C96712 | hypothetical prote |
| 22 | 35   | 54.7 | 225  | 2 | A97735 | hypothetical prote |
| 23 | 35   | 54.7 | 310  | 2 | AB0275 | arabinose operon r |
| 24 | 35   | 54.7 | 343  | 2 | AG1273 | N-acetylglutamate  |
| 25 | 35   | 54.7 | 343  | 2 | AH1636 | N-acetylglutamate  |
| 26 | 35   | 54.7 | 442  | 2 | T14353 | probable 4-hydroxy |
| 27 | 35   | 54.7 | 456  | 2 | E86903 | hypothetical prote |
| 28 | 35   | 54.7 | 469  | 2 | AC2794 | glutamine syntheta |
| 29 | 35   | 54.7 | 469  | 2 | B97573 | glutamine syntheta |
| 30 | 35   | 54.7 | 469  | 2 | AE3374 | glutamate-ammonia  |
| 31 | 35   | 54.7 | 473  | 2 | S75141 | glutamate-ammonia  |
| 32 | 35   | 54.7 | 572  | 2 | S55982 | asparagine synthas |
| 33 | 35   | 54.7 | 591  | 2 | A99444 | acylaminoacyl-pept |
| 34 | 35   | 54.7 | 657  | 2 | T24605 | hypothetical prote |
| 35 | 35   | 54.7 | 716  | 2 | A86181 | hypothetical prote |
| 36 | 35   | 54.7 | 939  | 2 | AE2275 | hypothetical prote |
| 37 | 35   | 54.7 | 1545 | 2 | B41859 | IgA-specific metal |
| 38 | 35   | 54.7 | 1802 | 2 | G71616 | hypothetical prote |
| 39 | 34.5 | 53.9 | 2076 | 2 | S15999 | fatty-acyl-CoA syn |
| 40 | 34   | 53.1 | 72   | 2 | G97134 | hypothetical prote |
| 41 | 34   | 53.1 | 75   | 2 | A86487 | unknown protein [i |
| 42 | 34   | 53.1 | 131  | 2 | G72653 | hypothetical prote |
| 43 | 34   | 53.1 | 169  | 2 | E97775 | hypothetical prote |
| 44 | 34   | 53.1 | 227  | 2 | T06362 | probable 2-oxoglut |
| 45 | 34   | 53.1 | 263  | 2 | A82069 | hypothetical prote |

#### ALIGNMENTS

##### RESULT 1

E83846

D-alanyl-D-alanine carboxypeptidase (penicillin-binding protein) BH1573

[imported] - *Bacillus halodurans* (strain C-125)

C;Species: *Bacillus halodurans*

C;Date: 01-Dec-2000 #sequence\_revision 01-Dec-2000 #text\_change 09-Jul-2004

C;Accession: E83846

R;Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; Hiramata, C.; Nakamura, Y.; Ogasawara, N.; Kuhara, S.; Horikoshi, K.

Nucleic Acids Res. 28, 4317-4331, 2000

A;Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and genomic sequence comparison with *Bacillus subtilis*.

A;Reference number: A83650; MUID:20512582; PMID:11058132

A;Accession: E83846

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-382 <STO>

A;Cross-references: UNIPROT:Q9KCJ8; GB:AP001512; GB:BA000004; NID:g10174030;

PIDN:BAB05292.1; GSPDB:GN00137

A;Experimental source: strain C-125  
C;Genetics:  
A;Gene: BH1573  
C;Superfamily: penicillin-binding protein 5

Query Match 64.1%; Score 41; DB 2; Length 382;  
Best Local Similarity 70.0%; Pred. No. 11;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GISNTVYANP 13  
|:||||: ||  
Db 148 GMSNTVFQNP 157

#### RESULT 2

T47905

hypothetical protein T20K12.30 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 20-Apr-2000 #sequence\_revision 20-Apr-2000 #text\_change 09-Jul-2004

C;Accession: T47905

R;De Haan, M.; Maarse, A.C.; Grivell, L.A.; Mewes, H.W.; Lemcke, K.; Mayer, K.F.X.; Quetier, F.; Salanoubat, M.

submitted to the Protein Sequence Database, January 2000

A;Reference number: Z24480

A;Accession: T47905

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-673 <DEH>

A;Cross-references: UNIPROT:Q9LE59; EMBL:AL137898

A;Experimental source: cultivar Columbia; BAC clone T20K12

C;Genetics:

A;Map position: 3

A;Introns: 51/1; 59/3; 79/3; 121/3; 143/1; 158/3; 270/2; 389/3; 582/3

A;Note: T20K12.30

C;Superfamily: Arabidopsis thaliana hypothetical protein T20K12.30

Query Match 60.9%; Score 39; DB 2; Length 673;  
Best Local Similarity 66.7%; Pred. No. 47;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
|| || | |||  
Db 461 SGSSNLKYRNPK 472

#### RESULT 3

C64208

hypothetical protein MG075 - Mycoplasma genitalium

C;Species: Mycoplasma genitalium

C;Date: 10-Nov-1995 #sequence\_revision 10-Nov-1995 #text\_change 09-Jul-2004

C;Accession: C64208

R;Fraser, C.M.; Gocayne, J.D.; White, O.; Adams, M.D.; Clayton, R.A.; Fleischmann, R.D.; Bult, C.J.; Kerlavage, A.R.; Sutton, G.; Kelley, J.M.; Fritchman, J.L.; Weidman, J.F.; Small, K.V.; Sandusky, M.; Fuhrmann, J.; Nguyen, D.; Utterback, T.R.; Saudek, D.M.; Phillips, C.A.; Merrick, J.M.; Tomb, J.F.; Dougherty, B.A.; Bott, K.F.; Hu, P.C.; Lucier, T.S.; Peterson, S.N.; Smith, H.O.; Hutchison III, C.A.; Venter, J.C.

Science 270, 397-403, 1995

A;Title: The minimal gene complement of Mycoplasma genitalium.

A;Reference number: A64200; MUID:96026346; PMID:7569993

A;Accession: C64208

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1024 <TIGR>

A;Cross-references: UNIPROT:P47321; GB:U39687; GB:L43967; NID:g1045744;

PID:g1045751; TIGR:MG075

A;Experimental source: strain G-37

C;Genetics:

A;Genetic code: SGC3

Query Match 60.9%; Score 39; DB 2; Length 1024;  
Best Local Similarity 60.0%; Pred. No. 73;  
Matches 6; Conservative 3; Mismatches 1; Indels 0; Gaps 0;

Qy 5 ISNTVYANPK 14  
: |||::|||  
Db 773 LQNTVFSNPK 782

#### RESULT 4

F82242

oligopeptide ABC transporter, ATP-binding protein VC1095 [imported] - *Vibrio cholerae* (strain N16961 serogroup O1)

C;Species: *Vibrio cholerae*

C;Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 09-Jul-2004

C;Accession: F82242

R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Haft, D.H.; Hickey, E.K.; Peterson, J.D.; Umayam, L.A.; Gill, S.R.; Nelson, K.E.; Read, T.D.; Tettelin, H.; Richardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P.; McDonald, L.; Utterback, T.; Fleishmann, R.D.; Nierman, W.C.; White, O.; Salzberg, S.L.; Smith, H.O.; Colwell, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000

A;Title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.

A;Reference number: A82035; MUID:20406833; PMID:10952301

A;Accession: F82242

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-336 <HEI>

A;Cross-references: UNIPROT:Q9KT10; GB:AE004190; GB:AEC03852; NID:g9655559;

PIDN:AAF94254.1; GSPDB:GN00126; TIGR:VC1095

A;Experimental source: serogroup O1; strain N16961; biotype El Tor

C;Genetics:

A;Gene: VC1095

A;Map position: 1

C;Superfamily: inner membrane protein malK; ATP-binding cassette homology

Query Match 59.4%; Score 38; DB 2; Length 336;  
Best Local Similarity 54.5%; Pred. No. 35;  
Matches 6; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14  
|:| ::|||

Db

250 GVSKELFANPK 260

RESULT 5

D69262

hypothetical protein AF0100 - *Archaeoglobus fulgidus*

C;Species: *Archaeoglobus fulgidus*

C;Date: 05-Dec-1997 #sequence\_revision 05-Dec-1997 #text\_change 09-Jul-2004

C;Accession: D69262

R;Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, R.J.; Gwinn, M.; Hickey, E.K.; Peterson, J.D.; Richardson, D.L.; Kerlavage, A.R.; Graham, D.E.; Kyrpides, N.C.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.F.; Dougherty, B.A.; McKenny, K.; Adams, M.D.; Loftus, B.; Peterson, S.; Reich, C.I.; McNeil, L.K.; Badger, J.H.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L.

Nature 390, 364-370, 1997

A;Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artiach, P.; Kaine, B.P.; Sykes, S.M.; Sadow, P.W.; D'Andrea, K.P.; Bowman, C.; Fujii, C.; Garland, S.A.; Mason, T.M.; Olsen, G.J.; Fraser, C.M.; Smith, H.O.; Woese, C.R.; Venter, J.C.

A;Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaeon *Archaeoglobus fulgidus*.

A;Reference number: A69250; MUID:98049343; PMID:9389475

A;Accession: D69262

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-93 <KLE>

A;Cross-references: UNIPROT:O30136; GB:AE001099; GB:AE000782; NID:g2689422;

PIDN:AAB91130.1; PID:g2650548; TIGR:AF0100

Query Match 57.8%; Score 37; DB 2; Length 93;  
Best Local Similarity 63.6%; Pred. No. 14;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14

||| | :|||

Db 9 GIENVVKSNNPK 19

RESULT 6

E84671

hypothetical protein At2g27320 [imported] - *Arabidopsis thaliana*

C;Species: *Arabidopsis thaliana* (mouse-ear cress)

C;Date: 02-Feb-2001 #sequence\_revision 02-Feb-2001 #text\_change 09-Jul-2004

C;Accession: E84671

R;Lin, X.; Kaul, S.; Rounsley, S.D.; Shea, T.P.; Benito, M.I.; Town, C.D.; Fujii, C.Y.; Mason, T.M.; Bowman, C.L.; Barnstead, M.E.; Feldblyum, T.V.; Buell, C.R.; Ketchum, K.A.; Lee, J.J.; Ronning, C.M.; Koo, H.; Moffat, K.S.; Cronin, L.A.; Shen, M.; VanAken, S.E.; Umayam, L.; Tallon, L.J.; Gill, J.E.; Adams, M.D.; Carrera, A.J.; Creasy, T.H.; Goodman, H.M.; Somerville, C.R.; Copenhaver, G.P.; Preuss, D.; Nierman, W.C.; White, O.; Eisen, J.A.; Salzberg, S.L.; Fraser, C.M.; Venter, J.C.

Nature 402, 761-768, 1999

A;Title: Sequence and analysis of chromosome 2 of the plant *Arabidopsis thaliana*.

A;Reference number: A84420; MUID:20083487; PMID:10617197

A;Accession: E84671

A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-339 <STO>  
A;Cross-references: UNIPROT:Q9XIN9; GB:AE002093; NID:g5306265; PIDN:AAD41997.1;  
GSPDB:GN00139  
C;Genetics:  
A;Gene: At2g27320  
A;Map position: 2

Query Match 57.8%; Score 37; DB 2; Length 339;  
Best Local Similarity 63.6%; Pred. No. 54;  
Matches 7; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14  
| : ||| |||  
Db 228 GGNETVYTNPk 238

#### RESULT 7

H81146

penicillin-binding protein NMB0877 [imported] - Neisseria meningitidis (strain MC58 serogroup B)

C;Species: Neisseria meningitidis

C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 09-Jul-2004

C;Accession: H81146

R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Ketchum, K.A.; Hood, D.W.; Peden, J.F.; Dodson, R.J.; Nelson, W.C.; Gwinn, M.L.; DeBoy, R.; Peterson, J.D.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Mason, T.; Ciecko, A.; Parksey, D.S.; Blair, E.; Cittone, H.; Clark, E.B.; Cotton, M.D.; Utterback, T.R.; Khouri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M.

Science 287, 1809-1815, 2000

A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.;

Rappuoli, R.; Venter, J.C.

A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.

A;Reference number: A81000; MUID:20175755; PMID:10710307

A;Accession: H81146

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-389 <TET>

A;Cross-references: UNIPROT:Q9JZW2; GB:AE002440; GB:AE002098; NID:g7226112;

PIDN:AAF41288.1; PID:g7226115; GSPDB:GN00119; TIGR:NMB0877

A;Experimental source: serogroup B, strain MC58

C;Genetics:

A;Gene: NMB0877

C;Superfamily: penicillin-binding protein 5

Query Match 57.8%; Score 37; DB 2; Length 389;  
Best Local Similarity 60.0%; Pred. No. 62;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GISNTVYANP 13  
| : ||| : ||  
Db 164 GMKNTVFKNP 173

RESULT 8

D81875

probable serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) NMA1095

[imported] - *Neisseria meningitidis* (strain Z2491 serogroup A)

C;Species: *Neisseria meningitidis*

C;Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 09-Jul-2004

C;Accession: D81875

R;Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morelli, G.; Basham, D.; Brown, D.; Chillingworth, T.; Davies, R.M.; Davis, P.; Devlin, K.; Feltwell, T.; Hamlin, N.; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream, M.A.; Rutherford, K.M.; Simmonds, M.; Skelton, J.; Whitehead, S.; Spratt, B.G.; Barrell, B.G. Nature 404, 502-506, 2000

A;Title: Complete DNA sequence of a serogroup A strain of *Neisseria meningitidis* Z2491.

A;Reference number: A81775; MUID:20222556; PMID:10761919

A;Accession: D81875

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-389 <PAR>

A;Cross-references: UNIPROT:Q9JUX6; GB:AL162755; GB:AL157959; NID:g7379742;

PIDN:CAB84358.1; PID:g7379790; GSPDB:GN00124; NMASP:NMA1095

A;Experimental source: serogroup A, strain Z2491

C;Genetics:

A;Gene: NMA1095

C;Superfamily: penicillin-binding protein 5

C;Keywords: hydrolase; serine carboxypeptidase

Query Match 57.8%; Score 37; DB 2; Length 389;  
Best Local Similarity 60.0%; Pred. No. 62;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GISNTVYANP 13  
|: |||: ||  
Db 164 GMKNTVFKNP 173

RESULT 9

S44287

pyruvate kinase, plastid - common tobacco

C;Species: *Nicotiana tabacum* (common tobacco)

C;Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 09-Jul-2004

C;Accession: S44287

R;Blakeley, S.; Gottlob-McHugh, S.; Wan, J.; Crews, L.; Miki, B.; Ko, K.; Dennis, D.

submitted to the EMBL Data Library, November 1993

A;Description: Molecular characterisation of plastid pyruvate kinase from castor and tobacco.

A;Reference number: S44286

A;Accession: S44287

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-562 <BLA>

A;Cross-references: UNIPROT:Q40546; EMBL:Z28374; NID:g482937; PIDN:CAA82223.1;

PID:g482938

C;Superfamily: pyruvate kinase

Query Match 57.8%; Score 37; DB 2; Length 562;  
Best Local Similarity 66.7%; Pred. No. 91;  
Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GISNTVYAN 12  
|::| |||  
Db 59 GVNNNVYAN 67

RESULT 10

F70383

organic solvent tolerance protein - Aquifex aeolicus

C;Species: Aquifex aeolicus

C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 09-Jul-2004

C;Accession: F70383

R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Overbeek, R.; Snead, M.A.; Keller, M.; Aujay, M.; Huber, R.; Feldman, R.A.; Short, J.M.; Olson, G.J.; Swanson, R.V.

Nature 392, 353-358, 1998

A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A;Reference number: A70300; MUID:98196666; PMID:9537320

A;Accession: F70383

A;Status: preliminary; nucleic acid sequence not shown; translation not shown.

A;Molecule type: DNA

A;Residues: 1-653 <AQF>

A;Cross-references: UNIPROT:O67097; GB:AE000716; NID:g2983478; PIDN:AAC07065.1; PID:g2983486; GB:AE000657

A;Experimental source: strain VF5

C;Genetics:

A;Gene: ostA

Query Match 57.8%; Score 37; DB 2; Length 653;  
Best Local Similarity 77.8%; Pred. No. 1.1e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GISNTVYAN 12  
||||:| |  
Db 556 GISNSVYKN 564

RESULT 11

E83031

conserved hypothetical protein PA4927 [imported] - Pseudomonas aeruginosa (strain PAO1)

C;Species: Pseudomonas aeruginosa

C;Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 09-Jul-2004

C;Accession: E83031

R;Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; Brinkman, F.S.L.; Hufnagle, W.O.; Kowalik, D.J.; Lagrou, M.; Garber, R.L.; Goltry, L.; Tolentino, E.; Westbrook-Wadman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, R.M.; Smith, K.A.; Spencer, D.H.; Wong, G.K.S.; Wu, Z.; Paulsen, I.T.; Reizer, J.; Saier, M.H.; Hancock, R.E.W.; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A;Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pathogen.  
 A;Reference number: A82950; MUID:20437337; PMID:10984043  
 A;Accession: E83031  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-830 <STO>  
 A;Cross-references: UNIPROT:Q9HUN7; GB:AE004905; GB:AE004091; NID:g9951195; PIDN:AAG08312.1; GSPDB:GN00131; PASP:PA4927  
 A;Experimental source: strain PA01  
 C;Genetics:  
 A;Gene: PA4927

Query Match 57.8%; Score 37; DB 2; Length 830;  
 Best Local Similarity 50.0%; Pred. No. 1.4e+02;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
 :|:: |||:  
 Db 58 NGVTYNVYADPK 69

# RESULT 12

G71518  
 hypothetical protein pmpB - *Chlamydia trachomatis* (serotype D, strain UW3/Cx).  
 C;Species: *Chlamydia trachomatis*  
 C;Date: 13-Sep-1998 #sequence\_revision 13-Sep-1998 #text\_change 08-Oct-1999  
 C;Accession: G71518  
 R;Stephens, R.S.; Kalman, S.; Lammel, C.J.; Fan, J.; Marathe, R.; Aravind, L.; Mitchell, W.P.; Olinger, L.; Tatusov, R.L.; Zhao, Q.; Koonin, E.V.; Davis, R.W. Science 282, 754-759, 1998  
 A;Title: Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*.  
 A;Reference number: A71570; MUID:99000809; PMID:9784136  
 A;Accession: G71518  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-1751 <ARN>  
 A;Cross-references: GB:AE001314; GB:AE001273; NID:g3328833; PIDN:AAC68010.1; PID:g3328841  
 A;Experimental source: serotype D, strain UW-3/Cx  
 C;Genetics:  
 A;Gene: pmpB

Query Match 57.8%; Score 37; DB 2; Length 1751;  
 Best Local Similarity 50.0%; Pred. No. 2.9e+02;  
 Matches 6; Conservative 4; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
 ||:|:: |||  
 Db 1303 SGVSSSIPTNPK 1314

# RESULT 13

C64378  
 hypothetical protein MJ0627 - *Methanococcus jannaschii*  
 C;Species: *Methanococcus jannaschii*



C;Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 09-Jul-2004  
 C;Accession: C64378  
 R;Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.;  
 Blake, J.A.; FitzGerald, L.M.; Clayton, R.A.; Gocayne, J.D.; Kerlavage, A.R.;  
 Dougherty, B.A.; Tomb, J.F.; Adams, M.D.; Reich, C.I.; Overbeek, R.; Kirkness,  
 E.F.; Weinstock, K.G.; Merrick, J.M.; Glodek, A.; Scott, J.L.; Geoghagen,  
 N.S.M.; Weidman, J.F.; Fuhrmann, J.L.; Nguyen, D.; Utterback, T.R.; Kelley,  
 J.M.; Peterson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.;  
 Hurst, M.A.  
 Science 273, 1058-1073, 1996  
 A;Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.;  
 Woese, C.R.; Venter, J.C.  
 A;Title: Complete genome sequence of the methanogenic archaeon, *Methanococcus*  
*jannaschii*.  
 A;Reference number: A64300; MUID:96337999; PMID:8688087  
 A;Accession: C64378  
 A;Status: preliminary; nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-117 <BUL>  
 A;Cross-references: UNIPROT:Q58044; GB:U67510; GB:L77117; NID:g1591325;  
 PIDN:AAB98626.1; PID:g1591338; TIGR:MJ0627  
 C;Genetics:  
 A;Map position: FOR555371-555724  
 C;Superfamily: *Methanococcus jannaschii* hypothetical protein MJ0627

Query Match 56.2%; Score 36; DB 2; Length 117;  
 Best Local Similarity 60.0%; Pred. No. 27;  
 Matches 6; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 5 ISNTVYANPK 14  
 | | : | ||  
 Db 93 IENKIYENPK 102

# RESULT 14

AC3366

serine-type D-Ala-D-Ala carboxypeptidase (EC 3.4.16.4) [imported] - *Brucella*  
*melitensis* (strain 16M)

C;Species: *Brucella melitensis*

C;Date: 01-Feb-2002 #sequence\_revision 01-Feb-2002 #text\_change 09-Jul-2004

C;Accession: AC3366

R;DelVecchio, V.G.; Kapatral, V.; Redkar, R.J.; Patra, G.; Mujer, C.; Los, T.;  
 Ivanova, N.; Anderson, I.; Bhattacharyya, A.; Lykidis, A.; Reznik, G.;  
 Jablonski, L.; Larsen, N.; D'Souza, M.; Bernal, A.; Mazur, M.; Goltsman, E.;  
 Selkov, E.; Elzer, P.H.; Hagius, S.; O'Callaghan, D.; Letesson, J.J.; Haselkorn,  
 R.; Kypides, N.; Overbeek, R.

Proc. Natl. Acad. Sci. U.S.A. 99, 443-448, 2002

A;Title: The genome sequence of the facultative intracellular pathogen *Brucella*  
*melitensis*.

A;Reference number: AD3252; PMID:11756688

A;Accession: AC3366

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-205 <KUR>

A;Cross-references: UNIPROT:Q8YH91; GB:AE008917; PIDN:AAL52094.1; PID:g17982866;  
 GSPDB:GN00190

A;Experimental source: strain 16M

C;Genetics:  
A;Gene: BMEI0913  
A;Map position: I  
C;Keywords: hydrolase; serine carboxypeptidase

Query Match 56.2%; Score 36; DB 2; Length 205;  
Best Local Similarity 50.0%; Pred. No. 49;  
Matches 5; Conservative 4; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GISNTVYANP 13  
|: :|::|||  
Db 161 GMKSTIFANP 170

RESULT 15

S43887

restriction endonuclease - *Neisseria lactamica*

C;Species: *Neisseria lactamica*

C;Date: 19-Mar-1997 #sequence\_revision 19-Mar-1997 #text\_change 09-Jul-2004

C;Accession: S43887

R;Lau, P.C.K.; Forghani, F.; Labbe, D.; Bergeron, H.; Brousseau, R.; Hoeltke, H.J.

Mol. Gen. Genet. 243, 24-31, 1994

A;Title: The NlaIV restriction and modification genes of *Neisseria lactamica* are flanked by leucine biosynthesis genes.

A;Reference number: S43885; MUID:94247353; PMID:8190068

A;Accession: S43887

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-243 <LAU>

A;Cross-references: UNIPROT:P50183; GB:U06074; NID:g476225; PIDN:AAA53238.1;  
PID:g476228

C;Superfamily: *Neisseria lactamica* restriction endonuclease

Query Match 56.2%; Score 36; DB 2; Length 243;  
Best Local Similarity 63.6%; Pred. No. 58;  
Matches 7; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

Qy 3 SGISNTVYANP 13  
| | ||| ||  
Db 193 SAIEETVYQNP 203

Search completed: January 31, 2005, 13:23:44.

Job time : 20.4545 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 12:56:50 ; Search time 106.591 Seconds  
(without alignments)  
75.572 Million cell updates/sec

Title: US-10-067-620-1

Perfect score: 64  
 Sequence: 1 XXSGISNTVYANPK 14  
 Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5  
 Searched: 1825181 seqs, 575374646 residues  
 Total number of hits satisfying chosen parameters: 1825181  
 Minimum DB seq length: 0  
 Maximum DB seq length: 2000000000  
 Post-processing: Minimum Match 0%  
 Maximum Match 100%  
 Listing first 45 summaries  
 Database : UniProt\_02:\*  
 1: uniprot\_sprot:\*  
 2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result<br>No. | %<br>Query |       | Length | DB | ID         | Description        |
|---------------|------------|-------|--------|----|------------|--------------------|
|               | Score      | Match |        |    |            |                    |
| 1.            | 43         | 67.2  | 666    | 2  | Q8IAZ6     | Q8iaz6 plasmodium  |
| 2             | 43         | 67.2  | 1564   | 2  | Q8I5W4     | Q8i5w4 plasmodium  |
| 3             | 42         | 65.6  | 540    | 2  | Q6HHP3     | Q6hhp3 bacillus th |
| 4             | 42         | 65.6  | 540    | 2  | Q81CC5     | Q81cc5 bacillus ce |
| 5             | 42         | 65.6  | 540    | 2  | Q81PF9     | Q81pf9 bacillus an |
| 6             | 42         | 65.6  | 540    | 2  | AAT31965   | Aat31965 bacillus  |
| 7             | 41         | 64.1  | 382    | 2  | Q9KCJ8     | Q9kcj8 bacillus ha |
| 8             | 41         | 64.1  | 1972   | 2  | Q6MS19     | Q6ms19 mycoplasma  |
| 9             | 41         | 64.1  | 1972   | 2  | CAE77572   | Cae77572 mycoplasm |
| 10            | 40         | 62.5  | 355    | 2  | Q7UPI1     | Q7upi1 rhodopirell |
| 11            | 40         | 62.5  | 620    | 2  | Q8WSK2     | Q8wsk2 trypanosoma |
| 12            | 39         | 60.9  | 106    | 2  | Q9BW72     | Q9bw72 homo sapien |
| 13            | 39         | 60.9  | 106    | 2  | Q9CQJ1     | Q9cqj1 m mus muscu |
| 14            | 39         | 60.9  | 260    | 2  | Q8RGL2     | Q8rgl2 fusobacteri |
| 15            | 39         | 60.9  | 276    | 2  | Q7RRX7     | Q7rrx7 plasmodium  |
| 16            | 39         | 60.9  | 278    | 2  | Q8IDP1     | Q8idp1 plasmodium  |
| 17            | 39         | 60.9  | 342    | 2  | Q7RK05     | Q7rk05 plasmodium  |
| 18            | 39         | 60.9  | 505    | 1  | GCSB_SULTO | Q972c0 sulfolobus  |
| 19            | 39         | 60.9  | 584    | 2  | Q8RGS8     | Q8rgs8 fusobacteri |
| 20            | 39         | 60.9  | 639    | 2  | Q94BZ8     | Q94bz8 arabidopsis |
| 21            | 39         | 60.9  | 655    | 2  | Q869S4     | Q869s4 dictyosteli |
| 22            | 39         | 60.9  | 673    | 2  | Q9LE59     | Q9le59 arabidopsis |
| 23            | 39         | 60.9  | 1024   | 1  | Y075_MYCGE | P47321 mycoplasma  |
| 24            | 39         | 60.9  | 1476   | 2  | Q8TFN3     | Q8tfn3 neurospora  |
| 25            | 39         | 60.9  | 1520   | 2  | Q7S8U6     | Q7s8u6 neurospora  |
| 26            | 38         | 59.4  | 67     | 2  | Q7P3N2     | Q7p3n2 fusobacteri |
| 27            | 38         | 59.4  | 204    | 2  | Q71AE7     | Q71ae7 mamestra co |

|    |    |      |      |   |            |                    |
|----|----|------|------|---|------------|--------------------|
| 28 | 38 | 59.4 | 204  | 2 | Q8QLG9     | Q8qlg9 mamestra co |
| 29 | 38 | 59.4 | 204  | 2 | AAQ11084   | Aaq11084 mamestra  |
| 30 | 38 | 59.4 | 315  | 2 | Q6CAL3     | Q6cal3 yarrowia li |
| 31 | 38 | 59.4 | 323  | 2 | Q9H635     | Q9h635 homo sapien |
| 32 | 38 | 59.4 | 333  | 2 | Q9F5R5     | Q9f5r5 vibrio chol |
| 33 | 38 | 59.4 | 336  | 2 | Q9KT10     | Q9kt10 vibrio chol |
| 34 | 38 | 59.4 | 356  | 1 | ACC1_MOUSE | Q9d8z1 mus musculu |
| 35 | 38 | 59.4 | 400  | 2 | Q6CFI7     | Q6cfi7 yarrowia li |
| 36 | 38 | 59.4 | 432  | 2 | Q874D8     | Q874d8 orpinomyces |
| 37 | 38 | 59.4 | 432  | 2 | Q874E0     | Q874e0 orpinomyces |
| 38 | 38 | 59.4 | 522  | 2 | Q8R585     | Q8r585 mus musculu |
| 39 | 38 | 59.4 | 690  | 2 | Q8BP56     | Q8bp56 m mus muscu |
| 40 | 38 | 59.4 | 690  | 2 | AAH56953   | Aah56953 mus muscu |
| 41 | 38 | 59.4 | 1542 | 2 | Q7SFA8     | Q7sfa8 neurospora  |
| 42 | 37 | 57.8 | 93   | 1 | Y100_ARCFU | O30136 archaeoglob |
| 43 | 37 | 57.8 | 204  | 2 | Q8JM88     | Q8jm88 mamestra co |
| 44 | 37 | 57.8 | 223  | 2 | Q6UTY3     | Q6uty3 cymbidium m |
| 45 | 37 | 57.8 | 223  | 2 | Q6UTY4     | Q6uty4 cymbidium m |

# ALIGNMENTS

## RESULT 1

### Q8IAZ6

ID Q8IAZ6 PRELIMINARY; PRT; 666 AA.  
AC Q8IAZ6;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Lipoamide dehydrogenase, putative (EC 1.8.1.4).  
GN Name=PF08\_0066;  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,  
RA Quail M., Barrell B.;  
RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
CC -!- COFACTOR: FAD (By similarity).  
CC -!- SIMILARITY: Belongs to the class-I pyridine nucleotide-disulfide  
CC oxidoreductase family.  
DR EMBL; AL844507; CAD51214.1; -.  
DR HSSP; Q94655; 1ONF.  
DR GO; GO:0005737; C:cytoplasm; IEA.  
DR GO; GO:0004148; F:dihydrolipoyl dehydrogenase activity; IEA.  
DR GO; GO:0015036; F:disulfide oxidoreductase activity; IEA.  
DR GO; GO:0006118; P:electron transport; IEA.  
DR InterPro; IPR001327; FAD\_pyr\_redox.  
DR InterPro; IPR001100; Pyr\_redox.  
DR InterPro; IPR004099; Pyr\_redox\_dim.  
DR Pfam; PF00070; Pyr\_redox; 1.  
DR Pfam; PF02852; Pyr\_redox\_dim; 1.  
DR PRINTS; PR00368; FADPNR.  
DR PRINTS; PR00411; PNDRDTASEI.  
DR ProDom; PD000139; FAD\_pyr\_redox; 1.

DR PROSITE; PS00076; PYRIDINE\_REDOX\_1; 1.  
KW FAD; Flavoprotein; Oxidoreductase; Redox-active center.  
SQ SEQUENCE 666 AA; 75587 MW; 1A876D357BBE3AEB CRC64;

Query Match 67.2%; Score 43; DB 2; Length 666;  
Best Local Similarity 72.7%; Pred. No. 48;  
Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14  
||:| || ||  
Db 52 GINNFVYINPK 62

## RESULT 2

Q8I5W4

ID Q8I5W4 PRELIMINARY; PRT; 1564 AA.  
AC Q8I5W4;  
DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN ORFNames=PFL0420w;  
OS Plasmodium falciparum (isolate 3D7).  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=36329;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22255705; PubMed=12368864;  
RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
RA Chan M.S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,  
RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
RA Fraser C.M., Barrell B.;  
RT "Genome sequence of the human malaria parasite Plasmodium  
RT falciparum.";  
RL Nature 419:498-511(2002).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Hyman R.W., Fung E., Conway A., Kurdi O., Mao J., Miranda M.,  
RA Nakao B., Rowley D., Tamaki T., Wang F., Davis R.W.;  
RL Submitted (JAN-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AE014845; AAN36173.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR GO; GO:0005279; F:amino acid-polyamine transporter activity; IEA.  
DR GO; GO:0006865; P:amino acid transport; IEA.  
DR InterPro; IPR002422; AA/rel\_permease2.  
KW Hypothetical protein.  
SQ SEQUENCE 1564 AA; 185930 MW; 086D9F972AE27786 CRC64;

Query Match 67.2%; Score 43; DB 2; Length 1564;  
Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
| |||:: |||  
Db 897 SNISNTLHINPK 908

RESULT 3

Q6HHP3

ID Q6HHP3 PRELIMINARY; PRT; 540 AA.  
AC Q6HHP3;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Oligopeptide ABC transporter, substrate-binding protein.  
GN Name=oppA; ORFNames=BT9727\_2608;  
OS Bacillus thuringiensis serovar konkukian str. 97-27.  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
OC Bacillus thuringiensis serovar konkukian.  
OX NCBI\_TaxID=281309;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=97-27;  
RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
RA Richardson P., Rubin E., Tice H.;  
RL Submitted (JUN-2004) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: Belongs to the bacterial extracellular solute-binding  
CC protein family 5.  
DR EMBL; AE017355; AAT61295.1; -.  
DR InterPro; IPR000914; SBP\_bac\_5.  
DR Pfam; PF00496; SBP\_bac\_5; 1.  
SQ SEQUENCE 540 AA; 62049 MW; FB4DD01F6A5EF202 CRC64;

Query Match 65.6%; Score 42; DB 2; Length 540;  
Best Local Similarity 58.3%; Pred. No. 59;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 3 SGISNTVYANPK 14  
| ::|| |||:  
Db 456 SSVNNTTEYANPE 467

RESULT 4

Q81CC5

ID Q81CC5 PRELIMINARY; PRT; 540 AA.  
AC Q81CC5;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Oligopeptide-binding protein oppA.  
GN ORFNames=BC2848;  
OS Bacillus cereus (strain ATCC 14579 / DSM 31).  
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
OX NCBI\_TaxID=226900;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=22608415; PubMed=12721630; DOI=10.1038/nature01582;  
RA Ivanova N., Sorokin A., Anderson I., Galleron N., Candelon B.,

RA Kapatral V., Bhattacharyya A., Reznik G., Mikhailova N., Lapidus A.,  
 RA Chu L., Mazur M., Goltsman E., Larsen N., D'Souza M., Walunas T.,  
 RA Grechkin Y., Pusch G., Haselkorn R., Fonstein M., Ehrlich S.D.,  
 RA Overbeek R., Kyrpides N.C.;  
 RT "Genome sequence of *Bacillus cereus* and comparative analysis with  
 RT *Bacillus anthracis*.";  
 RL Nature 423:87-91(2003).  
 CC -!- SIMILARITY: Belongs to the bacterial extracellular solute-binding  
 CC protein family 5.  
 DR EMBL; AE017007; AAP09798.1; -.  
 DR HSSP; P06202; 1JEV.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 SQ SEQUENCE 540 AA; 62034 MW; E49C7E79528C4DDA CRC64;

Query Match 65.6%; Score 42; DB 2; Length 540;  
 Best Local Similarity 58.3%; Pred. No. 59;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
 | ::|| ||||:  
 Db 456 SSVNNTTEYANPE 467

# RESULT 5

## Q81PF9

ID Q81PF9 PRELIMINARY; PRT; 540 AA.  
 AC Q81PF9; Q6HXM2; Q6KRP8;  
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Oligopeptide ABC transporter, oligopeptide-binding protein,  
 DE putative.  
 GN OrderedLocusNames=BA2848, BAS2657; ORFNames=GBAA2848;  
 OS *Bacillus anthracis*.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.  
 OX NCBI\_TaxID=1392;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate Porton;  
 RX MEDLINE=22608414; PubMed=12721629; DOI=10.1038/nature01586;  
 RA Read T.D., Peterson S.N., Tourasse N.J., Baillie L.W., Paulsen I.T.,  
 RA Nelson K.E., Tettelin H., Fouts D.E., Eisen J.A., Gill S.R.,  
 RA Holtzapple E.K., Okstad O.A., Helgason E., Rilstone J., Wu M.,  
 RA Kolonay J.F., Beanan M.J., Dodson R.J., Brinkac L.M., Gwinn M.L.,  
 RA DeBoy R.T., Madpu R., Daugherty S.C., Durkin A.S., Haft D.H.,  
 RA Nelson W.C., Peterson J.D., Pop M., Khouri H.M., Radune D.,  
 RA Benton J.L., Mahamoud Y., Jiang L., Hance I.R., Weidman J.F.,  
 RA Berry K.J., Plaut R.D., Wolf A.M., Watkins K.L., Nierman W.C.,  
 RA Hazen A., Cline R.T., Redmond C., Thwaite J.E., White O.,  
 RA Salzberg S.L., Thomason B., Friedlander A.M., Koehler T.M.,  
 RA Hanna P.C., Kolstoe A.-B., Fraser C.M.;  
 RT "The genome sequence of *Bacillus anthracis* Ames and comparison to  
 RT closely related bacteria.";  
 RL Nature 423:81-86(2003).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames / isolate 0581;  
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S.L.,  
 RA Fraser C.M.;  
 RT "Bacillus anthracis comparative genomics."  
 RL Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Sterne;  
 RA Brettin T.S., Bruce D., Challacombe J.F., Gilna P., Han C., Hill K.,  
 RA Hitchcock P., Jackson P., Keim P., Longmire J., Lucas S., Okinaka R.,  
 RA Richardson P., Rubin E., Tice H.;  
 RL Submitted (JAN-2004) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SIMILARITY: Belongs to the bacterial extracellular solute-binding  
 CC protein family 5.  
 DR EMBL; AE017033; AAP26678.1; -.  
 DR EMBL; AE017334; AAT31965.1; -.  
 DR EMBL; AE017225; AAT54967.1; -.  
 DR HSSP; P06202; 1JEV.  
 DR TIGR; BA2848; -.  
 DR GO; GO:0005215; F:transporter activity; IEA.  
 DR GO; GO:0006810; P:transport; IEA.  
 DR InterPro; IPR000914; SBP\_bac\_5.  
 DR Pfam; PF00496; SBP\_bac\_5; 1.  
 SQ SEQUENCE 540 AA; 62096 MW; 1CE17917316535AE CRC64;

Query Match 65.6%; Score 42; DB 2; Length 540;  
 Best Local Similarity 58.3%; Pred. No. 59;  
 Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
 | ::|| ||||:  
 Db 456 SSVNNTTEYANPE 467

# RESULT 6

AAT31965

ID AAT31965 PRELIMINARY; PRT; 540 AA.  
 AC AAT31965;  
 DT 01-JUN-2004 (TrEMBLrel. 27, Created)  
 DT 01-JUN-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 01-JUN-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Oligopeptide ABC transporter, oligopeptide-binding protein,  
 DE putative.  
 GN GBAA2848.  
 OS Bacillus anthracis str. Ames 0581.  
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus;  
 OC Bacillus cereus group; Bacillus anthracis.  
 OX NCBI\_TaxID=261594;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Ames 0581;  
 RA Ravel J., Rasko D.A., Shumway M.F., Jiang L., Cer R.Z., Federova N.B.,  
 RA Wilson M., Stanley S., Decker S., Read T.D., Salzberg S., Fraser C.M.;  
 RT "Bacillus anthracis comparative genomics.";



RL Submitted (MAY-2004) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AE017334; AAT31965.1; -.  
SQ SEQUENCE 540 AA; 62096 MW; 1CE17917316535AE CRC64;

Query Match 65.6%; Score 42; DB 2; Length 540;  
Best Local Similarity 58.3%; Pred. No. 59;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

QY 3 SGISNTVYANPK 14  
|::|||:|  
Db 456 SSVNNTTEYANPE 467

RESULT 7

Q9KQJ8

ID Q9KQJ8 PRELIMINARY; PRT; 382 AA.  
AC Q9KQJ8;  
DT 01-OCT-2000 (TrEMBLrel. 15, Created)  
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE D-alanyl-D-alanine carboxypeptidase (Penicillin-binding protein).  
GN Name=BH1573;  
OS Bacillus halodurans.  
CC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.  
QX NCBI\_TaxID=86665;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C-125;  
RX MEDLINE=20512582; PubMed=11058132;  
RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,  
RA Fuji F., Hiramata C., Nakamura Y., Ogasawara N., Kuhara S.,  
RA Horikoshi K.;  
RT "Complete genome sequence of the alkaliphilic bacterium Bacillus  
RT halodurans and genomic sequence comparison with Bacillus subtilis.";  
RL Nucleic Acids Res. 28:4317-4331(2000).  
DR EMBL; AP001512; BAB05292.1; -.  
DR PIR; E83846; E83846.  
DR HSSP; P39042; 1ES4.  
DR GO; GO:0004185; F:serine carboxypeptidase activity; IEA.  
DR GO; GO:0006508; P:proteolysis and peptidolysis; IEA.  
DR InterPro; IPR001967; Peptidase\_S11.  
DR Pfam; PF00768; Peptidase\_S11; 1.  
DR PRINTS; PR00725; DADACBPTASE1.  
KW Carboxypeptidase.  
SQ SEQUENCE 382 AA; 43119 MW; 5745572EB0EC1A4E CRC64;

Query Match 64.1%; Score 41; DB 2; Length 382;  
Best Local Similarity 70.0%; Pred. No. 64;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 GISNTVYANP 13  
|:||||:|  
Db 148 GMSNTVFQNP 157

RESULT 8

Q6MS19

ID Q6MS19 PRELIMINARY; PRT; 1972 AA.  
AC Q6MS19;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocusNames=MSC\_0963;  
OS Mycoplasma mycoides (subsp. mycoides SC).  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=44101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PG1;  
RX PubMed=14762060;  
RA Westberg J., Persson A., Holmberg A., Goesmann A., Lundeberg J.,  
RA Johansson K.-E., Pettersson B., Uhlen M.;  
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type  
RT strain PG1T, the causative agent of contagious bovine pleuropneumonia  
RT (CBPP).";  
RL Genome Res. 14:221-227(2004).  
DR EMBL; BX842645; CAE77572.1; -.  
DR InterPro; IPR000566; Lipocln\_cytFABP.  
DR InterPro; IPR006025; Pept\_M\_Zn\_BS.  
DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.  
DR PROSITE; PS00142; ZINC\_PROTEASE; UNKNOWN\_1.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 1972 AA; 226965 MW; 0F00F95B31043351 CRC64;

Query Match 64.1%; Score 41; DB 2; Length 1972;  
Best Local Similarity 77.8%; Pred. No. 3.3e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 6 SNTVYANPK 14  
|| :|||||  
Db 1186 SNLIYANPK 1194

# RESULT 9

CAE77572

ID CAE77572 PRELIMINARY; PRT; 1972 AA.  
AC CAE77572;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 13-APR-2004 (TrEMBLrel. 27, Last annotation update)  
DE Hypothetical protein.  
GN MSC\_0963.  
OS Mycoplasma mycoides (subsp. mycoides SC).  
OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.  
OX NCBI\_TaxID=44101;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=PG1;  
RX PubMed=14762060;  
RA Westberg J., Persson A., Holmberg A., Goesmann A., Lundeberg J.,  
RA Johansson K.-E., Pettersson B., Uhlen M.;  
RT "The genome sequence of Mycoplasma mycoides subsp. mycoides SC type  
RT strain PG1T, the causative agent of contagious bovine pleuropneumonia

RT (CBPP).";  
RL Genome Res. 14:221-227(2004).  
DR EMBL; BX842645; CAE77572.1; -.  
KW Hypothetical protein.  
SQ SEQUENCE 1972 AA; 226965 MW; 0F00F95B31043351 CRC64;

Query Match 64.1%; Score 41; DB 2; Length 1972;  
Best Local Similarity 77.8%; Pred. No. 3.3e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 6 SNTVYANPK 14  
||:|||||  
Db 1186 SNLIYANPK 1194

#### RESULT 10

##### Q7UPI1

ID Q7UPI1 PRELIMINARY; PRT; 355 AA.  
AC Q7UPI1;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Probable fimbrial protein.  
GN OrderedLocusNames=RB6920;  
OS Rhodopirellula baltica.  
OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;  
OC Planctomycetaceae; Pirellula.  
OX NCBI\_TaxID=117;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=1;  
RX MEDLINE=22735913; PubMed=12835416;  
RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,  
RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,  
RA Schlesner H., Amann R., Reinhardt R.;  
RT "Complete genome sequence of the marine planctomycete Pirellula sp.  
RT strain 1.";  
RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).  
DR EMBL; BX294145; CAD75081.1; -.  
DR InterPro; IPR011453; DUF1559.  
DR InterPro; IPR001120; Prok\_N\_methyl\_S.  
DR Pfam; PF07596; DUF1559; 1.  
DR PROSITE; PS00409; PROKAR\_NTER\_METHYL; 1.  
KW Complete proteome; Methylation.  
SQ SEQUENCE 355 AA; 39242 MW; E6F52F42EDB5F464 CRC64;

Query Match 62.5%; Score 40; DB 2; Length 355;  
Best Local Similarity 63.6%; Pred. No. 90;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14  
||:|:|  
Db 215 GINNSKYRNP 225

#### RESULT 11

##### Q8WSK2

ID Q8WSK2 PRELIMINARY; PRT; 620 AA.  
AC Q8WSK2;  
DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Phosphodiesterase.  
GN Name=PDE1;  
OS Trypanosoma brucei brucei.  
OC Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae; Trypanosoma.  
OX NCBI\_TaxID=5702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=EATRO 1125;  
RX PubMed=14728691;  
RA Kunz S., Kloeckner T., Essen L.O., Seebeck T., Boshart M.;  
RT "TbPDE1, a novel class I phosphodiesterase of Trypanosoma brucei.";  
RL Eur. J. Biochem. 271:637-647(2004).  
DR EMBL; AF253418; AAL58095.1; -.  
DR HSSP; Q07343; 1F0J.  
DR GO; GO:0004114; F:3',5'-cyclic-nucleotide phosphodiesterase a. . .; IEA.  
DR GO; GO:0003824; F:catalytic activity; IEA.  
DR GO; GO:0007165; P:signal transduction; IEA.  
DR InterPro; IPR003607; Met\_phos\_hydro.  
DR InterPro; IPR002073; PDEase.  
DR Pfam; PF00233; PDEase\_I; 1.  
DR PRINTS; PR00387; PDIESTERASE1.  
DR SMART; SM00471; HDC; 1.  
DR PROSITE; PS00126; PDEASE\_I; 1.  
SQ SEQUENCE 620 AA; 70337 MW; 08FF5F6891299801 CRC64;

Query Match 62.5%; Score 40; DB 2; Length 620;  
Best Local Similarity 77.8%; Pred. No. 1.6e+02;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 6 SNTVYANPK 14  
:| |||||  
Db 129 ANAVYANPK 137

# RESULT 12

Q9BW72

ID Q9BW72 PRELIMINARY; PRT; 106 AA.  
AC Q9BW72;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE Hypothetical protein MGC2198.  
GN Name=MGC2198;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Skin, and Muscle;  
RX MEDLINE=22388257; PubMed=12477932;  
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Skin;  
 RA Strausberg R.;  
 RL Submitted (NOV-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Muscle;  
 RA Strausberg R.;  
 RL Submitted (MAY-2001) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC000587; AAH00587.1; -.  
 DR EMBL; BC007502; AAH07502.1; -.  
 DR InterPro; IPR007667; HIG\_1\_N.  
 DR Pfam; PF04588; HIG\_1\_N; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 106 AA; 11528 MW; 205DE0AE39950053 CRC64;

Query Match 60.9%; Score 39; DB 2; Length 106;  
 Best Local Similarity 63.6%; Pred. No. 41;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 GISNTVYANPK 14  
 |:| ||| |:|  
 Db 23 GLSPTVYRNPE 33

# RESULT 13

Q9CQJ1

ID Q9CQJ1 PRELIMINARY; PRT; 106 AA.  
 AC Q9CQJ1;  
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)  
 DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
 DE Mus musculus adult male small intestine cDNA, RIKEN full-length  
 DE enriched library, clone:2010110M21 product:hypothetical protein, full  
 DE insert sequence (2010110M21Rik protein) (Mus musculus adult male  
 DE cerebellum cDNA, RIKEN full-length enriched library, clone:1500016M15  
 DE product:hypothetical protein, full insert sequence).

GN Name=2010110M21Rik;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Small intestine;  
 RX MEDLINE=99279253; PubMed=10349636;  
 RA Carninci P., Hayashizaki Y.;  
 RT "High-efficiency full-length cDNA cloning.";  
 RL Meth. Enzymol. 303:19-44(1999).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Small intestine;  
 RX MEDLINE=21085660; PubMed=11217851;  
 RA RIKEN FANTOM Consortium;  
 RT "Functional annotation of a full-length mouse cDNA collection.";  
 RL Nature 409:685-690(2001).  
 RN [3]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Small intestine;  
 RA The FANTOM Consortium,  
 RA the RIKEN Genome Exploration Research Group Phase I & II Team;  
 RT "Analysis of the mouse transcriptome based on functional annotation of  
 RT 60,770 full-length cDNAs.";  
 RL Nature 420:563-573(2002).  
 RN [4]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Small intestine;  
 RX MEDLINE=20499374; PubMed=11042159;  
 RA Carninci P., Shibata Y., Hayatsu N., Sugahara Y., Shibata K., Itoh M.,  
 RA Konno H., Okazaki Y., Muramatsu M., Hayashizaki Y.;  
 RT "Normalization and subtraction of cap-trapper-selected cDNAs to  
 RT prepare full-length cDNA libraries for rapid discovery of new genes.";  
 RL Genome Res. 10:1617-1630(2000).  
 RN [5]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Small intestine;  
 RX MEDLINE=20530913; PubMed=11076861;  
 RA Shibata K., Itoh M., Aizawa K., Nagaoka S., Sasaki N., Carninci P.,  
 RA Konno H., Akiyama J., Nishi K., Kitsunai T., Tashiro H., Itoh M.,  
 RA Sumi N., Ishii Y., Nakamura S., Hazama M., Nishine T., Harada A.,  
 RA Yamamoto R., Matsumoto H., Sakaguchi S., Ikegami T., Kashiwagi K.,  
 RA Fujiwake S., Inoue K., Togawa Y., Izawa M., Ohara E., Watahiki M.,  
 RA Yoneda Y., Ishikawa T., Ozawa K., Tanaka T., Matsuura S., Kawai J.,  
 RA Okazaki Y., Muramatsu M., Inoue Y., Kira A., Hayashizaki Y.;  
 RT "RIKEN integrated sequence analysis (RISA) system-384-format  
 RT sequencing pipeline with 384 multicapillary sequencer.";  
 RL Genome Res. 10:1757-1771(2000).  
 RN [6]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=C57BL/6J; TISSUE=Cerebellum, and Small intestine;  
 RA Adachi J., Aizawa K., Akahira S., Akimura T., Arai A., Aono H.,  
 RA Arakawa T., Bono H., Carninci P., Fukuda S., Fukunishi Y., Furuno M.,  
 RA Hanagaki T., Hara A., Hayatsu N., Hiramoto K., Hiraoka T., Hori F.,  
 RA Imotani K., Ishii Y., Itoh M., Izawa M., Kasukawa T., Kato H.,

RA Kawai J., Kojima Y., Konno H., Kouda M., Koya S., Kurihara C.,  
 RA Matsuyama T., Miyazaki A., Nishi K., Nomura K., Numazaki R., Ohno M.,  
 RA Okazaki Y., Okido T., Owa C., Saito H., Saito R., Sakai C., Sakai K.,  
 RA Sano H., Sasaki D., Shibata K., Shibata Y., Shinagawa A., Shiraki T.,  
 RA Sogabe Y., Suzuki H., Tagami M., Tagawa A., Takahashi F., Tanaka T.,  
 RA Tejima Y., Toya T., Yamamura T., Yasunishi A., Yoshida K., Yoshino M.,  
 RA Muramatsu M., Hayashizaki Y.;  
 RL Submitted (JUL-2000) to the EMBL/GenBank/DDBJ databases.  
 RN [7]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences."  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [8]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=FVB/N; TISSUE=Mammary tumor. C3;  
 RA Strausberg R.;  
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AK008387; BAB25641.1; -.  
 DR EMBL; BC021471; AAH21471.1; -.  
 DR EMBL; AK005269; BAB23921.1; -.  
 DR MGD; MGI:1914294; 2010110M21Rik.  
 DR InterPro; IPR007667; HIG\_1\_N.  
 DR Pfam; PF04588; HIG\_1\_N; 1.  
 KW Hypothetical protein.  
 SQ SEQUENCE 106 AA; 11368 MW; 862E4D9015500A1D CRC64;

Query Match 60.9%; Score 39; DB 2; Length 106;  
 Best Local Similarity 63.6%; Pred. No. 41;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GISNTVYANPK 14  
 | | |||:|:  
 Db 23 GFSPTVYSNPE 33

RESULT 14  
 Q8RGL2

ID Q8RGL2 PRELIMINARY; PRT; 260 AA.  
AC Q8RGL2;  
DT 01-JUN-2002 (TrEMBLrel. 21, Created)  
DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Lipoprotein.  
GN OrderedLocusNames=FN0279;  
OS Fusobacterium nucleatum (subsp. nucleatum).  
OC Bacteria; Fusobacteria; Fusobacteriales; Fusobacteriaceae;  
OC Fusobacterium.  
OX NCBI\_TaxID=76856;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=ATCC 25586;  
RX MEDLINE=21886394; PubMed=11889109;  
RA Kapatral V., Anderson I., Ivanova N., Reznik G., Los T., Lykidis A.,  
RA Bhattacharyya A., Bartman A., Gardner W., Grichkin G., Zhu L.,  
RA Vasieva O., Chu L., Kogan Y., Chaga O., Goltsman E., Bernal A.,  
RA Larsen N., D'Souza M., Walunas T., Pusch G., Haselkorn R.,  
RA Fonstein M., Kyrpides N.C., Overbeek R.;  
RT "Genome sequence and analysis of the oral bacterium Fusobacterium  
RT nucleatum strain ATCC 25586.";  
RL J. Bacteriol. 184:2005-2018(2002).  
DR EMBL; AE010540; AAL94485.1; -.  
DR GO; GO:0016020; C:membrane; IEA.  
DR InterPro; IPR007428; VacJ.  
DR Pfam; PF04333; VacJ; 1.  
DR PRINTS; PR01805; VACJLIPOPROT.  
KW Complete proteome; Lipoprotein.  
SQ SEQUENCE 260 AA; 29659 MW; 2C72D15882883350 CRC64;

Query Match 60.9%; Score 39; DB 2; Length 260;  
Best Local Similarity 72.7%; Pred. No. 1e+02;  
Matches 3; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 3 SGISNTVYANP 13  
| || |||||  
Db 33 SEASNVVYANP 43

#### RESULT 15

##### Q7RRX7

ID Q7RRX7 PRELIMINARY; PRT; 276 AA.  
AC Q7RRX7;  
DT 01-MAR-2004 (TrEMBLrel. 26, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE Putative ubiquitin-conjugating enzyme 16.  
GN Name=PY00590;  
OS Plasmodium yoelii yoelii.  
OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
OX NCBI\_TaxID=73239;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=17XNL;  
RX PubMed=12368865;  
RA Carlton J.M., Angiuoli S.V., Suh B.B., Kooij T.W., Pertea M.,



RA Silva J.C., Ermolaeva M.D., Allen J.E., Selengut J.D., Koo H.L.,  
 RA Peterson J.D., Pop M., Kosack D.S., Shumway M.F., Bidwell S.L.,  
 RA Shallom S.J., van Aken S.E., Riedmuller S.B., Feldblyum T.V.,  
 RA Cho J.K., Quackenbush J., Sedegah M., Shoaibi A., Cummings L.M.,  
 RA Florens L., Yates F.R. III, Raine J.D., Sinden R.E., Harris M.A.,  
 RA Cunningham D.A., Preiser P.R., Bergman L.W., Vaidya A.B.,  
 RA van Lin L.H., Janse C.J., Waters A.P., Smith H.O., White O.R.,  
 RA Salzberg S.L., Venter J.C., Fraser C.M., Hoffman S.L., Gardner M.J.,  
 RA Carucci D.J.;  
 RT "Genome sequence and comparative analysis of the model rodent malaria  
 RT parasite *Plasmodium yoelii yoelii*.";  
 RL Nature 419:512-519(2002).  
 CC -!- FUNCTION: Catalyzes the covalent attachment of ubiquitin to other  
 CC proteins (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + ubiquitin + protein lysine = AMP +  
 CC diphosphate + protein N-ubiquityllysine.  
 CC -!- PATHWAY: Ubiquitin conjugation; second step.  
 CC -!- MISCELLANEOUS: A cysteine residue is required for ubiquitin-  
 CC thiolester formation (By similarity).  
 CC -!- SIMILARITY: Belongs to the ubiquitin-conjugating enzyme family.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AABL01000159; EAA17068.1; -.  
 DR GO; GO:0004840; F:ubiquitin conjugating enzyme activity; IEA.  
 DR GO; GO:0006512; P:ubiquitin cycle; IEA.  
 DR InterPro; IPR000608; UBQ\_conjugat.  
 DR Pfam; PF00179; UQ\_con; 1.  
 DR ProDom; PD000461; UBQ\_conjugat; 1.  
 DR PROSITE; PS50127; UBIQUITIN\_CONJUGAT\_2; 1.  
 KW Ligase; Ubl conjugation pathway.  
 SQ SEQUENCE 276 AA; 31727 MW; 26795275462E74B1 CRC64;

Query Match 60.9%; Score 39; DB 2; Length 276;  
 Best Local Similarity 66.7%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 GISNTVYAN 12  
 |: ||: |||  
 Db 170 GLENTIYAN 178

Search completed: January 31, 2005, 13:22:40  
 Job time : 109.591 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 12:55:49 ; Search time 105.091 Seconds  
(without alignments)  
54.616 Million cell updates/sec

Title: US-10-067-620-6  
Perfect score: 90  
Sequence: 1 LLDNLHQQTTPPDGFGR 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | % Query |       | Length | DB | ID       | Description        |
|------------|---------|-------|--------|----|----------|--------------------|
|            | Score   | Match |        |    |          |                    |
| 1          | 90      | 100.0 | 16     | 5  | ABB81973 | Abb81973 30 kDa ra |
| 2          | 49      | 54.4  | 20     | 2  | AAW71732 | Aaw71732 Rabbit 3- |
| 3          | 49      | 54.4  | 35     | 3  | AAB13216 | Aab13216 Human PDK |
| 4          | 49      | 54.4  | 113    | 3  | AAB06191 | Aab06191 Mammalian |
| 5          | 49      | 54.4  | 285    | 6  | ABR57461 | Abr57461 AGC famil |
| 6          | 49      | 54.4  | 289    | 8  | ADJ38865 | Adj38865 PDK1 amin |
| 7          | 49      | 54.4  | 319    | 6  | ABU04709 | Abu04709 Human exp |
| 8          | 49      | 54.4  | 319    | 6  | ABU04720 | Abu04720 Human exp |
| 9          | 49      | 54.4  | 335    | 4  | AAB99847 | Aab99847 AGC prote |

|    |    |      |     |   |          |          |           |
|----|----|------|-----|---|----------|----------|-----------|
| 10 | 49 | 54.4 | 335 | 8 | ADJ38895 | Adj38895 | PDK1 amin |
| 11 | 49 | 54.4 | 468 | 6 | ABU04719 | Abu04719 | Human exp |
| 12 | 49 | 54.4 | 468 | 6 | ABU04705 | Abu04705 | Human exp |
| 13 | 49 | 54.4 | 506 | 2 | AAy05780 | Aay05780 | Human pro |
| 14 | 49 | 54.4 | 506 | 6 | ABU04715 | Abu04715 | Human exp |
| 15 | 49 | 54.4 | 535 | 4 | AAB99823 | Aab99823 | AGC prote |
| 16 | 49 | 54.4 | 535 | 6 | ABU04713 | Abu04713 | Human exp |
| 17 | 49 | 54.4 | 556 | 2 | AAW71738 | Aaw71738 | Human 3-p |
| 18 | 49 | 54.4 | 556 | 2 | AAy27055 | Aay27055 | Human pro |
| 19 | 49 | 54.4 | 556 | 2 | AAy05779 | Aay05779 | Human pro |
| 20 | 49 | 54.4 | 556 | 3 | AAB28445 | Aab28445 | Human PDK |
| 21 | 49 | 54.4 | 556 | 3 | AAB28446 | Aab28446 | Human PDK |
| 22 | 49 | 54.4 | 556 | 3 | AAy94735 | Aay94735 | Phosphoin |
| 23 | 49 | 54.4 | 556 | 6 | ABO07176 | Abo07176 | Human p53 |
| 24 | 49 | 54.4 | 556 | 6 | ABU04708 | Abu04708 | Human exp |
| 25 | 49 | 54.4 | 556 | 6 | ABU04718 | Abu04718 | Human exp |
| 26 | 49 | 54.4 | 556 | 6 | ABU04712 | Abu04712 | Human exp |
| 27 | 49 | 54.4 | 556 | 6 | ABU04716 | Abu04716 | Human exp |
| 28 | 49 | 54.4 | 556 | 6 | ABU04711 | Abu04711 | Human exp |
| 29 | 49 | 54.4 | 556 | 6 | ABU04706 | Abu04706 | Human exp |
| 30 | 49 | 54.4 | 556 | 6 | ABU04714 | Abu04714 | Human exp |
| 31 | 49 | 54.4 | 556 | 6 | ABU04707 | Abu04707 | Human exp |
| 32 | 49 | 54.4 | 556 | 6 | ABU04717 | Abu04717 | Human exp |
| 33 | 49 | 54.4 | 556 | 7 | ABM79012 | Abm79012 | Human pho |
| 34 | 49 | 54.4 | 556 | 7 | ADD44919 | Add44919 | Human Pro |
| 35 | 49 | 54.4 | 556 | 7 | ADD44915 | Add44915 | Human Pro |
| 36 | 49 | 54.4 | 556 | 7 | ADD89983 | Add89983 | Human can |
| 37 | 49 | 54.4 | 556 | 8 | ADI36055 | Adi36055 | Human pho |
| 38 | 49 | 54.4 | 556 | 8 | ADO15485 | Ado15485 | Human PDP |
| 39 | 49 | 54.4 | 556 | 8 | ADQ19234 | Adq19234 | Human sof |
| 40 | 49 | 54.4 | 559 | 7 | ADD44917 | Add44917 | Rat Prote |
| 41 | 49 | 54.4 | 559 | 7 | ADD44913 | Add44913 | Rat Prote |
| 42 | 45 | 50.0 | 407 | 4 | AAU43451 | Aau43451 | Propionib |
| 43 | 45 | 50.0 | 407 | 6 | ABM39970 | Abm39970 | Propionib |
| 44 | 44 | 48.9 | 82  | 4 | AAG77402 | Aag77402 | Human col |
| 45 | 44 | 48.9 | 249 | 4 | ABB68467 | Abb68467 | Drosophil |

#### ALIGNMENTS

##### RESULT 1

ABB81973

ID ABB81973 standard; peptide; 16 AA.

XX

AC ABB81973;

XX

DT 25-NOV-2002 (first entry)

XX

DE 30 kDa ragweed pollen allergen tryptic peptide 6.

XX

KW Ragweed; pollen; allergen; Ambt 7; glycoprotein; antiallergic;  
KW immunotherapy; disulphide protein.

XX

OS Ambrosia elatior.

XX

PN WO200263012-A2.

XX  
 PD 15-AUG-2002.  
 XX  
 PF 04-FEB-2002; 2002WO-US003346.  
 XX  
 PR 05-FEB-2001; 2001US-0266686P.  
 XX  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Buchanan BB, Del Val G, Frick OL;  
 XX  
 DR WPI; 2002-657539/70.  
 XX  
 PT New ragweed pollen allergens, useful in allergy testing and immunotherapy  
 PT regimens, particularly for treating sensitivity to pollen or pollen  
 PT allergy (e.g. anaphylaxis, or symptoms of hives or asthma) in a mammal,  
 PT especially a human.  
 XX  
 PS Claim 1; Page 53; 70pp; English.  
 XX  
 CC The invention relates to an isolated pollen allergen purified from  
 CC ragweed pollen, substantially free of any other pollen proteins, or a  
 CC protein that is an antigenic fragment of a pollen allergen Ambt 7. The  
 CC allergen is characterized by the following physiochemical and biological  
 CC properties: (a) being contained in pollen extracts; (b) a glycoprotein;  
 CC (c) a sulphhydryl group containing protein; (d) a molecular weight of  
 CC about 30 kDa as determined by SDS-polyacrylamide gel electrophoresis; and  
 CC (e) possessing allergen activity. The pollen allergen, or antigenic  
 CC protein fragment of the pollen allergen Ambt 7, or composition is useful  
 CC for treating sensitivity to pollen or pollen allergy in a mammal. This  
 CC allergy includes anaphylaxis or atopy, which includes the symptoms of hay  
 CC fever, asthma or hives. The allergen is also useful in allergy testing  
 CC and immunotherapy regimens. Sequences ABB81968-978 represent tryptic  
 CC peptide fragments of the 30 kDa ragweed complete pollen extract  
 CC disulphide protein allergen  
 XX  
 SQ Sequence 16 AA;

Query Match 100.0%; Score 90; DB 5; Length 16;  
 Best Local Similarity 100.0%; Pred. No. 2.9e-08;  
 Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLDNLHQQTTPPDGFGR 16  
 |||||  
 Db 1 LLDNLHQQTTPPDGFGR 16

## RESULT 2

AAW71732

ID AAW71732 standard; peptide; 20 AA.

XX

AC AAW71732;

XX

DT 10-DEC-1998 (first entry)

XX

DE Rabbit 3-phosphoinositide dependent protein kinase peptide #4.

XX

KW Protein kinase B-alpha; 3-phosphoinositide-dependent protein kinase;  
KW diabetes; cancer; cell proliferation; phosphorylation.

XX

OS Oryctolagus cuniculus.

XX

PN WO9841638-A1.

XX

PD 24-SEP-1998.

XX

PF 16-MAR-1998; 98WO-GB000777.

XX

PR 17-MAR-1997; 97GB-00005462.

PR 19-JUN-1997; 97GB-00012826.

PR 15-AUG-1997; 97GB-00017253.

PR 03-OCT-1997; 97US-00943667.

XX

PA (MEDI-) MEDICAL RES COUNCIL.

XX

PI Alessi DR;

XX

DR WPI; 1998-531572/45.

XX

PT New isolated 3-phosphoinositide-dependent protein kinase - which  
PT phosphorylates and activates protein kinase B-alpha, used to develop  
PT products for treating diabetes or cancers or for enhancing cell  
PT proliferation.

XX

PS Example 2; Page 57; 120pp; English.

XX

CC A pure 3-phosphoinositide-dependent protein kinase (3PDPK) that  
CC phosphorylates and activates PK B-alpha has been isolated. The present  
CC sequence represents a rabbit 3-phosphoinositide dependent protein kinase  
CC peptide. Products from the present invention (e.g. 3PDPK, nucleotide  
CC sequence encoding 3PDPK, antibodies against 3PDPK) can be used to  
CC identify compounds which modulate the PK activity e.g. for treating  
CC diabetes or cancers or for enhancing cell proliferation in the  
CC regeneration of nerves or in wound healing

XX

SQ Sequence 20 AA;

Query Match 54.4%; Score 49; DB 2; Length 20;

Best Local Similarity 88.9%; Pred. No. 0.42;

Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTTP 11

:|||||||

Db 11 ENLHQQTTP 19

RESULT 3

AAB13216

ID AAB13216 standard; peptide; 35 AA.

XX

AC AAB13216;

XX

DT 11-JAN-2001 (first entry)

XX

DE Human PDK domain #4.  
 XX  
 KW Human; PDK domain; pdk-1; AKT kinase; daf-18; insulin signalling pathway;  
 KW daf-2; age-1; insulin receptor; PI 3-kinase; PKB kinase;  
 KW PTEN lipid phosphatase; antidiabetic; anorectic; obesity; diabetes.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200033068-A1.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 02-DEC-1999; 99WO-US028529.  
 XX  
 PR 03-DEC-1998; 98US-00205658.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 XX  
 PI Ruvkun G, Ogg S;  
 XX  
 DR WPI; 2000-423022/36.  
 XX  
 PT Diagnosing and treating obesity and impaired glucose tolerance using  
 PT modulators of daf-18 expression and/or activity.  
 XX  
 PS Disclosure; Page 363; 402pp; English.  
 XX  
 CC The present sequence is a human PDK domain which shows homology to pdk-1  
 CC from *Caenorhabditis elegans*. A number of *C. elegans* genes have been  
 CC identified as homologues of genes in the mammalian insulin signalling  
 CC pathway. The *C. elegans* age-1 gene encodes a homologue of the mammalian  
 CC PI 3-kinase whilst daf-2 encodes a homologue of the mammalian insulin  
 CC receptor. The *C. elegans* AKT kinase and PKB kinase act downstream of daf-  
 CC 2 and age-1, just as their mammalian homologues act downstream of insulin  
 CC signalling. The *C. elegans* PTEN lipid phosphatase homologue, DAF-18, has  
 CC been found to act upstream of AKT in the pathway. This discovery has  
 CC enabled mammalian PTEN action to be mapped to the insulin signalling  
 CC pathway. Conserved DAF motifs can be used to design probes to identify  
 CC mammalian DAF homologues and thus to identify individuals with a  
 CC predisposition toward the development of glucose intolerance conditions,  
 CC such as obesity and diabetes  
 XX  
 SQ Sequence 35 AA;

Query Match 54.4%; Score 49; DB 3; Length 35;  
 Best Local Similarity 88.9%; Pred. No. 0.77;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
 :|||||||  
 Db 13 ENLHQQTTP 21

#### RESULT 4

AAB06191

ID AAB06191 standard; protein; 113 AA.

XX

AC AAB06191;  
 XX  
 DT 11-JAN-2001 (first entry)  
 XX  
 DE Mammalian PDK domain #2.  
 XX  
 KW Human; mouse; PDK domain; pdk-1; AKT kinase; daf-18;  
 KW insulin signalling pathway; daf-2; age-1; insulin receptor; PI 3-kinase;  
 KW PKB kinase; PTEN lipid phosphatase; antidiabetic; anorectic; obesity;  
 KW diabetes.  
 XX  
 OS Homo sapiens.  
 OS Mus musculus.  
 XX  
 PN WO200033068-A1.  
 XX  
 PD 08-JUN-2000.  
 XX  
 PF 02-DEC-1999; 99WO-US028529.  
 XX  
 PR 03-DEC-1998; 98US-00205658.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 XX  
 PI Ruvkun G, Ogg S;  
 XX  
 DR WPI; 2000-423022/36.  
 XX  
 PT Diagnosing and treating obesity and impaired glucose tolerance using  
 PT modulators of daf-18 expression and/or activity.  
 XX  
 PS Disclosure; Page 357; 402pp; English.  
 XX  
 CC The present sequence is a domain of mammalian PDK which shows homology to  
 CC pdk-1 from Caenorhabditis elegans. A number of C. elegans genes have been  
 CC identified as homologues of genes in the mammalian insulin signalling  
 CC pathway. The C. elegans age-1 gene encodes a homologue of the mammalian  
 CC PI 3-kinase whilst daf-2 encodes a homologue of the mammalian insulin  
 CC receptor. The C. elegans AKT kinase and PKB kinase act downstream of daf-  
 CC 2 and age-1, just as their mammalian homologues act downstream of insulin  
 CC signalling. The C. elegans PTEN lipid phosphatase homologue, DAF-18, has  
 CC been found to act upstream of AKT in the pathway. This discovery has  
 CC enabled mammalian PTEN action to be mapped to the insulin signalling  
 CC pathway. Conserved DAF motifs can be used to design probes to identify  
 CC mammalian DAF homologues and thus to identify individuals with a  
 CC predisposition toward the development of glucose intolerance conditions,  
 CC such as obesity and diabetes  
 XX  
 SQ Sequence 113 AA;

Query Match 54.4%; Score 49; DB 3; Length 113;  
 Best Local Similarity 88.9%; Pred. No. 2.8;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTTP 11  
 :|||||||  
 Db 86 ENLHQQTTP 94

RESULT 5

ABR57461

ID ABR57461 standard; protein; 285 AA.

XX

AC ABR57461;

XX

DT 15-SEP-2003 (first entry)

XX

DE AGC family protein kinase protein PDK1.

XX

KW Protein kinase B; PKB/Akt; enzyme; crystal structure; drug discovery;  
KW protein co-ordinate data; cytostatic; antidiabetic; vasotropic; PKB;  
KW nootropic; neuroprotective; gene therapy; protein kinase B beta; PKBbeta;  
KW structural analysis; cancer; diabetes; erectile dysfunction;  
KW neurodegeneration.

XX

OS Unidentified.

XX

PN WO2003016516-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-GB003735.

XX

PR 14-AUG-2001; 2001GB-00019860.

PR 01-MAY-2002; 2002GB-00009985.

XX

PA (NOVS ) NOVARTIS FORSCHUNGSSTIFTUNG ZWEIGNIEDERL.

PA (CANC-) CANCER RES INST.

XX

PI Barford D, Yang J, Hemmings BA, Cron PD;

XX

DR WPI; 2003-268328/26.

XX

PT New crystal of protein kinase B beta, useful for activating protein  
PT kinases, e.g. AGC kinases, comprises three-dimensional atomic coordinates  
PT or a tetragonal space group.

XX

PS Disclosure; Fig 4; 284pp; English.

XX

CC The present invention describes a crystal of protein kinase B beta  
CC (PKBbeta) comprising (I), where (I) comprises: (a) a tetragonal space  
CC group P4-1-2-1-2 and unit cell dimensions of: a = 149.33 plus or minus  
CC 0.5 Angstrom, b = 149.33 plus or minus 0.5 Angstrom, c = 39.77 plus or  
CC minus 0.5 Angstrom; a = 148.40 plus or minus 0.5 Angstrom, b = 148.40  
CC plus or minus 0.5 Angstrom, c = 38.55 plus or minus 0.5 Angstrom; a =  
CC 149.70 plus or minus 0.5 Angstrom, b = 149.70 plus or minus 0.5 Angstrom,  
CC c = 39.19 plus or minus 0.5 Angstrom; or a = 149.52 plus or minus 0.5  
CC Angstrom, b = 149.52 plus or minus 0.5 Angstrom, c = 39.06 plus or minus  
CC 0.5 Angstrom; or (b) the three-dimensional atomic coordinates listed in  
CC the specification. (I) has cytostatic, antidiabetic, vasotropic,  
CC nootropic and neuroprotective activities, and can be used in gene  
CC therapy. The crystal of PKBbeta, and methods from the present invention,  
CC are useful in activating protein kinases, particularly AGC kinases, for  
CC identifying modulators of protein kinase activity, and for structural



CC analysis of other protein kinases. The crystal may also be used in  
CC manufacturing a medicament for treating cancers, diabetes, erectile  
CC dysfunction or neurodegeneration. The present sequence represents an AGC  
CC family protein kinase which is given in the exemplification of the  
CC present invention

XX

SQ Sequence 285 AA;

Query Match 54.4%; Score 49; DB 6; Length 285;  
Best Local Similarity 88.9%; Pred. No. 7.8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTTP 11  
:|||||||  
Db 273 ENLHQQTTP 281

# RESULT 6

ADJ38865

ID ADJ38865 standard; protein; 289 AA.

XX

AC ADJ38865;

XX

DT 06-MAY-2004 (first entry)

XX

DE PDK1 amino acid sequence.

XX

KW phosphoinositide dependent protein kinase 1; PDK1; molecular modelling;  
KW protein kinase; catalytic domain; enzyme; hydrophobic pocket;  
KW insulin signalling pathway; signalling; crystalline form;  
KW protein co-ordinate data; three-dimensional structure; antifungal;  
KW antidiabetic; cardiatic; cytostatic; cerebroprotective; vasotropic;  
KW anorectic; protein kinase modulator; cancer; diabetes; obesity;  
KW apoptosis inhibition; ischaemia disease; stroke; myocardial infarction;  
KW neural injury.

XX

OS Unidentified.

XX

PN WO2003104481-A2.

XX

PD 18-DEC-2003.

XX

PF 09-JUN-2003; 2003WO-GB002509.

XX

PR 08-JUN-2002; 2002GB-00013186.

XX

PA (UYDU-) UNIV DUNDEE.

XX

PI Alessi D, Biondi R, Komander D, Van AD;

XX

DR WPI; 2004-062373/06.

XX

PT Selecting/designing compound for modulating activity of phosphoinositide  
PT dependent protein kinase 1 by using molecular modelling to select/design  
PT compound predicted to interact with protein kinase catalytic domain.

XX

PS Example 1; Fig 3; 383pp; English.

XX

CC The present invention describes a method (M1) for selecting or designing  
CC a compound for modulating the activity of phosphoinositide dependent  
CC protein kinase 1 (PDK1) comprising using molecular modelling means to  
CC select or design a compound that is predicted to interact with the  
CC protein kinase catalytic domain of PDK1, and selecting a compound that is  
CC predicted to interact with the protein kinase catalytic domain. Also  
CC described: (1) selecting or designing (M2) a compound for modulating the  
CC activity of a hydrophobic pocket (PIF binding pocket)-containing protein  
CC kinase having a hydrophobic pocket in the position equivalent to the  
CC hydrophobic pocket of human PDK1 that is defined by residues including  
CC Lys115, Ile118, Ile119, Val124, Val127 and/or Leu155 of full-length human  
CC PDK1 and further having a phosphate binding pocket in the position  
CC equivalent to the phosphate binding pocket of human PDK1 that is defined  
CC by residues including Lys76, Arg131, Thr148 and/or Gln150; (2) assessing  
CC (M3) the activation state of a structure for a protein kinase; (3) a  
CC mutated protein kinase (I); (4) a polynucleotide (II) encoding (I); (5) a  
CC host cell (III) comprising (II); (6) identifying (M4) a compound that  
CC modulates the protein kinase activity of a protein kinase (e.g., PDK1);  
CC (7) an antibody (IV) reactive with the phosphate binding pocket of PDK1  
CC or (I) or an antibody reactive with PDK1 or (I) but not with the protein  
CC kinase mutated at the phosphate binding site, or vice versa; (8) a  
CC compound (V) identified or identifiable by (M1) or (M3); (9) use of (V),  
CC (I), (II) in medicine; (10) use of (V), (I), (II) for the manufacture of  
CC a medicament for the treatment of a patient in need of modulation of  
CC signalling by a protein kinase as defined, for example PDK1, SGK, PKB, or  
CC p70 S6 kinase, for example insulin signalling pathway and/or  
CC PDK1/PDK2/SGK/PKB/p70 S6 kinase/PRK2/PKC signalling; and (11) a  
CC crystalline form (VI) of polypeptide as defined in (M1). (I) has  
CC antifungal, antidiabetic, cardiant, cytostatic, cerebroprotective,  
CC vasotropic and anorectic activities, and can be used as a modulator of  
CC protein kinase. (V) is useful for modulating the ability of protein  
CC kinase to phosphorylate different substrates; e.g., different naturally  
CC occurring polypeptides, to different extents. (V) inhibits or increases  
CC the activity of protein kinase. The protein structures e.g., the co-  
CC ordinates as provided in the specification are useful for designing  
CC reagent useful in drug designing assays or characterisation of protein  
CC kinase activity or regulation. (V) capable of producing the activity of  
CC PKC, e.g., PKC beta, PRK1 or PRK2, PDK1, PKB, SGK or p70 S6 kinase, is  
CC useful in treating cancer. (V) capable of increasing the activity of  
CC PDK1, PKB, SGK or p70 S6 kinase is useful in treating diabetes or obesity  
CC or may be useful in inhibiting apoptosis, thus useful in treating  
CC diseases in which apoptosis is involved e.g., mechanical (including heat)  
CC tissue injury or ischaemia disease such as stroke, myocardial infarction  
CC and neural injury. (V) is useful as an antifungal agent. The present  
CC sequence is used in the exemplification of the present invention.

XX

SQ Sequence 289 AA;

Query Match 54.4%; Score 49; DB 8; Length 289;  
Best Local Similarity 88.9%; Pred. No. 8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTPP 11  
:|||||||  
Db 278 ENLHQQTPP 286

RESULT 7

ABU04709

ID ABU04709 standard; protein; 319 AA.

XX

AC ABU04709;

XX

DT 29-JAN-2003 (first entry)

XX

DE Human expressed protein tag (EPT) #1375.

XX

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
KW protease; protease inhibitor; transporter; cytoskeletal protein;  
KW receptor; transcription factor; cancer; MHC;  
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200278524-A2.

XX

PD 10-OCT-2002.

XX

PF 28-MAR-2002; 2002WO-US009671.

XX

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

XX

PA (ZYCO-) ZYCOS INC.

XX

PI Chicz RM, Tomlinson AJ, Urban RG;

XX

DR WPI; 2003-040607/03.

XX

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for  
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
PT leukemia.

XX

PS Example 2; SEQ ID NO 1375; 134pp; English.

XX

CC The invention describes a purified polypeptide, which comprises a  
CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
CC transporter, cytoskeletal protein, receptor or transcription factor. The  
CC polypeptide is useful as an immunogenic composition for eliciting in a  
CC mammal an immunogenic response directed against any of the purified  
CC polypeptide. The purified polypeptide, or the antibody that binds to this  
CC polypeptide, is useful for treating cancer. The polypeptide is also  
CC useful for identifying compounds that binds to a naturally processed  
CC class I or class II MHC-binding polypeptide. The polypeptides and  
CC polynucleotides are particularly useful for treating or preventing  
CC myeloma, colon cancer, gastric cancer; adenocarcinoma, sarcoma, melanoma,  
CC lymphoma or leukaemia. These are also useful for screening agents for

CC treating the above mentioned diseases. This sequence represents an  
CC expressed protein tag (EPT) isolated from human tissue for translational  
CC profiling. Note: This sequence does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 319 AA;

Query Match 54.4%; Score 49; DB 6; Length 319;  
Best Local Similarity 88.9%; Pred. No. 8.9;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTPP 11  
:|||||||  
Db 111 ENLHQQTPP 119

#### RESULT 8

ABU04720

ID ABU04720 standard; protein; 319 AA.

XX

AC ABU04720;

XX

DT 29-JAN-2003 (first entry)

XX

DE Human expressed protein tag (EPT) #1386.

XX

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
KW protease; protease inhibitor; transporter; cytoskeletal protein;  
KW receptor; transcription factor; cancer; MHC;  
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200278524-A2.

XX

PD 10-OCT-2002.

XX

PF 28-MAR-2002; 2002WO-US009671.

XX

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

XX

PA (ZYCO-) ZYCOS INC.

XX

PI Chicz RM, Tomlinson AJ, Urban RG;

XX

DR WPI; 2003-040607/03.

XX

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for  
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or

PT leukemia.  
 XX  
 PS Example 2; SEQ ID NO 1386; 134pp; English.  
 XX  
 CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukaemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 319 AA;

Query Match 54.4%; Score 49; DB 6; Length 319;  
 Best Local Similarity 88.9%; Pred. No. 8.9;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTTP 11  
 :|||||||  
 Db 111 ENLHQQTTP 119

# RESULT 9

AAB99847

ID AAB99847 standard; protein; 335 AA.  
 XX  
 AC AAB99847;  
 XX  
 DT 20-SEP-2001 (first entry)  
 XX  
 DE AGC protein kinase family member PDK1 protein sequence.  
 XX  
 KW Protein kinase; identification; hydrophobic pocket; interacting; cancer;  
 KW diabetes; inhibition; apoptosis; tissue injury; ischaemic injury; stroke.  
 XX  
 OS Homo sapiens.  
 OS Synthetic.  
 XX  
 PN WO200144497-A2.  
 XX  
 PD 21-JUN-2001.  
 XX  
 PF 04-DEC-2000; 2000WO-GB004598.  
 XX  
 PR 02-DEC-1999; 99US-0168559P.  
 XX

PA (UYDU-) UNIV DUNDEE.  
 XX  
 PI Alessi D, Biondi R;  
 XX  
 DR WPI; 2001-390252/41.  
 XX  
 PT Identifying modulators of protein kinase (PK) activity, useful in  
 PT developing drugs for treating cancer or diabetes, by measuring the  
 PT ability of the compound to modulate or mimic the interaction of PK with  
 PT interacting polypeptides.  
 XX  
 PS Disclosure; Fig 16; 180pp; English.  
 XX  
 CC The present invention describes a method for identifying a compound that  
 CC modulates protein kinase activity. The method comprises measuring the  
 CC ability of the compound to inhibit, promote or mimic the interaction of a  
 CC hydrophobic pocket-containing protein kinase with an interacting  
 CC polypeptide. The interacting polypeptide interacts with the hydrophobic  
 CC pocket of the protein kinase and/or comprises the amino acid sequence  
 CC Phe/Tyr-Xaa-Xaa-Phe/Tyr (I). The method is useful in screening assays for  
 CC developing pharmaceutical compounds or drugs. Compounds, polypeptides or  
 CC polynucleotides from the present invention are useful in medicine,  
 CC particularly in the manufacture of a medicament for treating a patient in  
 CC need of modulation of signalling by a hydrophobic pocket-containing  
 CC protein kinase. Specifically, the patient has cancer or diabetes or is in  
 CC need of inhibition of apoptosis, e.g. a patient suffering from tissue  
 CC injury or ischaemic injury, including stroke. The compound or composition  
 CC is also useful for inhibiting the degree or rate of phosphorylation by  
 CC the protein kinase. The interacting polypeptide or compound is useful in  
 CC methods of stabilising a hydrophobic pocket- containing protein kinase,  
 CC where the protein kinase is exposed to the compound or polypeptide.  
 CC AAB99786 to AAB99847 represent amino acid sequences, and AAH44210 and  
 CC AAH44211 represent oligonucleotide sequences, used in the exemplification  
 CC of the present invention  
 XX  
 SQ Sequence 335 AA;

Query Match 54.4%; Score 49; DB 4; Length 335;  
 Best Local Similarity 88.9%; Pred. No. 9.4;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTPP 11  
 :|||||||  
 Db 254 ENLHQQTPP 262

# RESULT 10

ADJ38895

ID ADJ38895 standard; protein; 335 AA.

XX

AC ADJ38895;

XX

DT 06-MAY-2004 (first entry)

XX

DE PDK1 amino acid sequence.

XX

KW phosphoinositide dependent protein kinase 1; PDK1; molecular modelling;

KW protein kinase; catalytic domain; enzyme; hydrophobic pocket;  
KW insulin signalling pathway; signalling; crystalline form;  
KW protein co-ordinate data; three-dimensional structure; antifungal;  
KW antidiabetic; cardiant; cytostatic; cerebroprotective; vasotropic;  
KW anorectic; protein kinase modulator; cancer; diabetes; obesity;  
KW apoptosis inhibition; ischaemia disease; stroke; myocardial infarction;  
KW neural injury.

XX  
OS Unidentified.

XX  
PN WO2003104481-A2.

XX  
PD 18-DEC-2003.

XX  
PF 09-JUN-2003; 2003WO-GB002509.

XX  
PR 08-JUN-2002; 2002GB-00013186.

XX  
PA (UYDU-) UNIV DUNDEE.

XX  
PI Alessi D, Biondi R, Komander D, Van AD;

XX  
DR WPI; 2004-062373/06.

XX  
PT Selecting/designing compound for modulating activity of phosphoinositide  
PT dependent protein kinase 1 by using molecular modelling to select/design  
PT compound predicted to interact with protein kinase catalytic domain.

XX  
PS Disclosure; Fig 7; 383pp; English.

XX  
CC The present invention describes a method (M1) for selecting or designing  
CC a compound for modulating the activity of phosphoinositide dependent  
CC protein kinase 1 (PDK1) comprising using molecular modelling means to  
CC select or design a compound that is predicted to interact with the  
CC protein kinase catalytic domain of PDK1, and selecting a compound that is  
CC predicted to interact with the protein kinase catalytic domain. Also  
CC described: (1) selecting or designing (M2) a compound for modulating the  
CC activity of a hydrophobic pocket (PIF binding pocket)-containing protein  
CC kinase having a hydrophobic pocket in the position equivalent to the  
CC hydrophobic pocket of human PDK1 that is defined by residues including  
CC Lys115, Ile118, Ile119, Val124, Val127 and/or Leu155 of full-length human  
CC PDK1 and further having a phosphate binding pocket in the position  
CC equivalent to the phosphate binding pocket of human PDK1 that is defined  
CC by residues including Lys76, Arg131, Thr148 and/or Gln150; (2) assessing  
CC (M3) the activation state of a structure for a protein kinase; (3) a  
CC mutated protein kinase (I); (4) a polynucleotide (II) encoding (I); (5) a  
CC host cell (III) comprising (II); (6) identifying (M4) a compound that  
CC modulates the protein kinase activity of a protein kinase (e.g., PDK1);  
CC (7) an antibody (IV) reactive with the phosphate binding pocket of PDK1  
CC or (I) or an antibody reactive with PDK1 or (I) but not with the protein  
CC kinase mutated at the phosphate binding site, or vice versa; (8) a  
CC compound (V) identified or identifiable by (M1) or (M3); (9) use of (V),  
CC (I), (II) in medicine; (10) use of (V), (I), (II) for the manufacture of  
CC a medicament for the treatment of a patient in need of modulation of  
CC signalling by a protein kinase as defined, for example PDK1, SGK, PKB, or  
CC p70 S6 kinase, for example insulin signalling pathway and/or  
CC PDK1/PDK2/SGK/PKB/p70 S6 kinase/PRK2/PKC signalling; and (11) a

CC crystalline form (VI) of polypeptide as defined in (M1). (I) has  
 CC antifungal, antidiabetic, cardiant, cytostatic, cerebroprotective,  
 CC vasotropic and anorectic activities, and can be used as a modulator of  
 CC protein kinase. (V) is useful for modulating the ability of protein  
 CC kinase to phosphorylate different substrates, e.g., different naturally  
 CC occurring polypeptides, to different extents. (V) inhibits or increases  
 CC the activity of protein kinase. The protein structures e.g., the co-  
 CC ordinates as provided in the specification are useful for designing  
 CC reagent useful in drug designing assays or characterisation of protein  
 CC kinase activity or regulation. (V) capable of producing the activity of  
 CC PKC, e.g., PKC beta, PRK1 or PRK2, PDK1, PKB, SGK or p70 S6 kinase, is  
 CC useful in treating cancer. (V) capable of increasing the activity of  
 CC PDK1, PKB, SGK or p70 S6 kinase is useful in treating diabetes or obesity  
 CC or may be useful in inhibiting apoptosis, thus useful in treating  
 CC diseases in which apoptosis is involved e.g., mechanical (including heat)  
 CC tissue injury or ischaemia disease such as stroke, myocardial infarction  
 CC and neural injury. (V) is useful as an antifungal agent. The present  
 CC sequence is used in the exemplification of the present invention.  
 XX  
 SQ Sequence 335 AA;

Query Match 54.4%; Score 49; DB 8; Length 335;  
 Best Local Similarity 88.9%; Pred. No. 9.4;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTPP 11  
 :|||||||  
 Db 254 ENLHQQTPP 262

# RESULT 11

ABU04719

ID ABU04719 standard; protein; 468 AA.

XX

AC ABU04719;

XX

DT 29-JAN-2003 (first entry)

XX

DE Human expressed protein tag (EPT) #1385.

XX

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
 KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200278524-A2.

XX

PD 10-OCT-2002.

XX

PF 28-MAR-2002; 2002WO-US009671.

XX

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.



PR 01-OCT-2001; 2001US-0326370P.  
PR 04-DEC-2001; 2001US-0336780P.  
PR 20-FEB-2002; 2002US-0358985P.

XX

PA (ZYCO-) ZYCOS INC.

XX

PI Chicz RM, Tomlinson AJ, Urban RG;

XX

DR WPI; 2003-040607/03.

XX

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for  
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
PT leukemia.

XX

PS Example 2; SEQ ID NO 1385; 134pp; English.

XX

CC The invention describes a purified polypeptide, which comprises a  
CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
CC transporter, cytoskeletal protein, receptor or transcription factor. The  
CC polypeptide is useful as an immunogenic composition for eliciting in a  
CC mammal an immunogenic response directed against any of the purified  
CC polypeptide. The purified polypeptide, or the antibody that binds to this  
CC polypeptide, is useful for treating cancer. The polypeptide is also  
CC useful for identifying compounds that binds to a naturally processed  
CC class I or class II MHC-binding polypeptide. The polypeptides and  
CC polynucleotides are particularly useful for treating or preventing  
CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
CC lymphoma or leukaemia. These are also useful for screening agents for  
CC treating the above mentioned diseases. This sequence represents an  
CC expressed protein tag (EPT) isolated from human tissue for translational  
CC profiling. Note: This sequence does not appear in the printed  
CC specification but was obtained in electronic format directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 468 AA;

Query Match 54.4%; Score 49; DB 6; Length 468;  
Best Local Similarity 88.9%; Pred. No. 14;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTPP 11  
:|||||||  
Db 260 ENLHQQTPP 268

RESULT 12

ABU04705

ID ABU04705 standard; protein; 468 AA.

XX

AC ABU04705;

XX

DT 29-JAN-2003 (first entry)

XX

DE Human expressed protein tag (EPT) #1371.

XX

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;

KW protease; protease inhibitor; transporter; cytoskeletal protein;  
 KW receptor; transcription factor; cancer; MHC;  
 KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
 KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200278524-A2.  
 XX  
 PD 10-OCT-2002.  
 XX  
 PF 28-MAR-2002; 2002WO-US009671.  
 XX  
 PR 28-MAR-2001; 2001US-0279495P.  
 PR 21-MAY-2001; 2001US-0292544P.  
 PR 08-AUG-2001; 2001US-0310801P.  
 PR 01-OCT-2001; 2001US-0326370P.  
 PR 04-DEC-2001; 2001US-0336780P.  
 PR 20-FEB-2002; 2002US-0358985P.  
 XX  
 PA (ZYCO-) ZYCOS INC.  
 XX  
 PI Chicz RM, Tomlinson AJ, Urban RG;  
 XX  
 DR WPI; 2003-040607/03.  
 XX  
 PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
 PT cytoskeletal proteins, receptors or transcription factors), useful for  
 PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or  
 PT leukemia.  
 XX  
 PS Example 2; SEQ ID NO 1371; 134pp; English.  
 XX  
 CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukaemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 468 AA;

Query Match 54.4%; Score 49; DB 6; Length 468;  
 Best Local Similarity 88.9%; Pred. No. 14;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTPP 11  
:|||||||  
Db 260 ENLHQQTPP 268

RESULT 13

AAAY05780

ID AAY05780 standard; protein; 506 AA.

XX

AC AAY05780;

XX

DT 02-AUG-1999 (first entry)

XX

DE Human protein kinase B kinase.

XX

KW Protein kinase B kinase; PKB kinase; inhibitor; assay; cytostatic;

KW cell proliferation; cancer; therapy; signal transduction; human.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Domain 31. .303

FT /note= "protein kinase domain"

FT Domain 367. .494

FT /note= "PH domain"

XX

PN WO9916887-A2.

XX

PD 08-APR-1999.

XX

PF 17-SEP-1998; 98WO-US019412.

XX

PR 26-SEP-1997; 97US-0060190P.

XX

PA (ONYX-) ONYX PHARM INC.

XX

PI Stephens L, Hawkings P, Stokoe D;

XX

DR WPI; 1999-263699/22.

DR N-PSDB; AAX25486.

XX

PT Protein kinase B kinase nucleotide sequence and product.

XX

PS Example 1; Fig 3; 38pp; English.

XX

CC The present sequence represents a 55 kDa protein kinase B (PKB) kinase  
CC that activates PKB in the signal transduction pathway of  
CC phosphatidylinositol-3,4,5-trisphosphate (PIP3). The sequence is  
CC predicted from EST clones and cDNAs isolated from a human U937 cell  
CC library. The following are claimed: (1) an isolated nucleic acid molecule  
CC containing a nucleotide sequence which encodes PKB kinase activity; (2) a  
CC nucleotide sequence encoding a chimeric protein comprising the nucleic  
CC acid molecule fused to a second nucleotide sequence encoding a  
CC heterologous protein; (3) an expression vector; (4) a host cell; (5) an  
CC antibody that immunospecifically binds to PKB kinase; (6) a method for  
CC diagnosing disease in a mammal by detecting a PKB kinase gene mutation in  
CC the mammal's genome; (7) a method for screening compounds for treatment

CC of cell growth disorders utilising activated PKB kinase; (8) activation  
CC of PKB kinase by incubation in solution with PIP3; (9) compounds  
CC identified in (7); and (10) an isolated PKB kinase. PKB is involved in  
CC regulating cell growth, hence PKB kinase inhibitors can be used to treat  
CC disease involving unwanted cell growth, including cancer

XX

SQ Sequence 506 AA;

Query Match 54.4%; Score 49; DB 2; Length 506;  
Best Local Similarity 88.9%; Pred. No. 15;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTTP 11  
:|||||||  
Db 298 ENLHQQTTP 306

#### RESULT 14

ABU04715

ID ABU04715 standard; protein; 506 AA.

XX

AC ABU04715;

XX

DT 29-JAN-2003 (first entry)

XX

DE Human expressed protein tag (EPT) #1381.

XX

KW Translational profiling; expressed protein tag; EPT; kinase; phosphatase;  
KW protease; protease inhibitor; transporter; cytoskeletal protein;  
KW receptor; transcription factor; cancer; MHC;  
KW major histocompatibility complex; myeloma; colon cancer; gastric cancer;  
KW adenocarcinoma; sarcoma; melanoma; lymphoma; leukaemia.

XX

OS Homo sapiens.

XX

PN WO200278524-A2.

XX

PD 10-OCT-2002.

XX

PF 28-MAR-2002; 2002WO-US009671.

XX

PR 28-MAR-2001; 2001US-0279495P.

PR 21-MAY-2001; 2001US-0292544P.

PR 08-AUG-2001; 2001US-0310801P.

PR 01-OCT-2001; 2001US-0326370P.

PR 04-DEC-2001; 2001US-0336780P.

PR 20-FEB-2002; 2002US-0358985P.

XX

PA (ZYCO-) ZYCOS INC.

XX

PI Chicz RM, Tomlinson AJ, Urban RG;

XX

DR WPI; 2003-040607/03.

XX

PT New polypeptides (e.g. kinases, phosphatases, proteases, transporters,  
PT cytoskeletal proteins, receptors or transcription factors), useful for  
PT treating cancer, e.g. colon cancer, gastric cancer, sarcoma, lymphoma or

PT leukemia.

XX

PS Example 2; SEQ ID NO 1381; 134pp; English.

XX

CC The invention describes a purified polypeptide, which comprises a  
 CC fragment of a kinase, phosphatase, protease, protease inhibitor,  
 CC transporter, cytoskeletal protein, receptor or transcription factor. The  
 CC polypeptide is useful as an immunogenic composition for eliciting in a  
 CC mammal an immunogenic response directed against any of the purified  
 CC polypeptide. The purified polypeptide, or the antibody that binds to this  
 CC polypeptide, is useful for treating cancer. The polypeptide is also  
 CC useful for identifying compounds that binds to a naturally processed  
 CC class I or class II MHC-binding polypeptide. The polypeptides and  
 CC polynucleotides are particularly useful for treating or preventing  
 CC myeloma, colon cancer, gastric cancer, adenocarcinoma, sarcoma, melanoma,  
 CC lymphoma or leukaemia. These are also useful for screening agents for  
 CC treating the above mentioned diseases. This sequence represents an  
 CC expressed protein tag (EPT) isolated from human tissue for translational  
 CC profiling. Note: This sequence does not appear in the printed  
 CC specification but was obtained in electronic format directly from WIPO at  
 CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 506 AA;

Query Match 54.4%; Score 49; DB 6; Length 506;  
 Best Local Similarity 88.9%; Pred. No. 15;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
 :|||||||  
 Db 298 ENLHQQTTP 306

# RESULT 15

AAB99823

ID AAB99823 standard; protein; 535 AA.

XX

AC AAB99823;

XX

DT 20-SEP-2001 (first entry)

XX

DE AGC protein kinase family member PDK1 protein sequence.

XX

KW Protein kinase; identification; hydrophobic pocket; interacting; cancer;  
 KW diabetes; inhibition; apoptosis; tissue injury; ischaemic injury; stroke.

XX

OS Homo sapiens.

OS Synthetic.

XX

PN WO200144497-A2.

XX

PD 21-JUN-2001.

XX

PF 04-DEC-2000; 2000WO-GB004598.

XX

PR 02-DEC-1999; 99US-0168559P.

XX

PA (UYDU-) UNIV DUNDEE.

XX

PI Alessi D, Biondi R;

XX

DR WPI; 2001-390252/41.

XX

PT Identifying modulators of protein kinase (PK) activity, useful in  
PT developing drugs for treating cancer or diabetes, by measuring the  
PT ability of the compound to modulate or mimic the interaction of PK with  
PT interacting polypeptides.

XX

PS Disclosure; Fig 15; 180pp; English.

XX

CC The present invention describes a method for identifying a compound that  
CC modulates protein kinase activity. The method comprises measuring the  
CC ability of the compound to inhibit, promote or mimic the interaction of a  
CC hydrophobic pocket-containing protein kinase with an interacting  
CC polypeptide. The interacting polypeptide interacts with the hydrophobic  
CC pocket of the protein kinase and/or comprises the amino acid sequence  
CC Phe/Tyr-Xaa-Xaa-Phe/Tyr (I). The method is useful in screening assays for  
CC developing pharmaceutical compounds or drugs. Compounds, polypeptides or  
CC polynucleotides from the present invention are useful in medicine,  
CC particularly in the manufacture of a medicament for treating a patient in  
CC need of modulation of signalling by a hydrophobic pocket-containing  
CC protein kinase. Specifically, the patient has cancer or diabetes or is in  
CC need of inhibition of apoptosis, e.g. a patient suffering from tissue  
CC injury or ischaemic injury, including stroke. The compound or composition  
CC is also useful for inhibiting the degree or rate of phosphorylation by  
CC the protein kinase. The interacting polypeptide or compound is useful in  
CC methods of stabilising a hydrophobic pocket- containing protein kinase,  
CC where the protein kinase is exposed to the compound or polypeptide.  
CC AAB99786 to AAB99847 represent amino acid sequences, and AAH44210 and  
CC AAH44211 represent oligonucleotide sequences, used in the exemplification  
CC of the present invention

XX

SQ Sequence 535 AA;

Query Match 54.4%; Score 49; DB 4; Length 535;  
Best Local Similarity 88.9%; Pred. No. 16;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTPP 11  
:|||||||  
Db 327 ENLHQQTPP 335

Search completed: January 31, 2005, 13:17:02  
Job time : 107.091 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 13:08:40 ; Search time 28.3636 Seconds  
(without alignments)

37.410 Million cell updates/sec

Title: US-10-067-620-6  
Perfect score: 90  
Sequence: 1 LLDNLHQQTTPPDGFR 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID | Description                            |
|------------|-------|---------|--------------|-------|----------------------------------------|
| 1          | 49    | 54.4    | 20           | 4     | US-08-943-667-11 Sequence 11, Appl     |
| 2          | 49    | 54.4    | 556          | 2     | US-09-016-000-4 Sequence 4, Appli      |
| 3          | 49    | 54.4    | 556          | 4     | US-09-156-793D-2 Sequence 2, Appli     |
| 4          | 49    | 54.4    | 556          | 4     | US-08-943-667-1 Sequence 1, Appli      |
| 5          | 44    | 48.9    | 141          | 4     | US-09-270-767-37903 Sequence 37903, A  |
| 6          | 44    | 48.9    | 141          | 4     | US-09-270-767-53120 Sequence 53120, A  |
| 7          | 44    | 48.9    | 152          | 4     | US-09-270-767-56695 Sequence 56695, A  |
| 8          | 44    | 48.9    | 226          | 4     | US-09-270-767-31989 Sequence 31989, A  |
| 9          | 44    | 48.9    | 226          | 4     | US-09-270-767-47206 Sequence 47206, A  |
| 10         | 44    | 48.9    | 389          | 4     | US-09-270-767-41474 Sequence 41474, A  |
| 11         | 44    | 48.9    | 390          | 4     | US-09-252-991A-33044 Sequence 33044, A |
| 12         | 42.5  | 47.2    | 289          | 4     | US-09-107-532A-6795 Sequence 6795, Ap  |
| 13         | 42    | 46.7    | 589          | 4     | US-09-543-681A-4194 Sequence 4194, Ap  |
| 14         | 42    | 46.7    | 1092         | 4     | US-09-543-681A-7058 Sequence 7058, Ap  |
| 15         | 41    | 45.6    | 31           | 4     | US-09-270-767-61771 Sequence 61771, A  |
| 16         | 41    | 45.6    | 117          | 4     | US-09-621-976-4623 Sequence 4623, Ap   |
| 17         | 41    | 45.6    | 186          | 1     | US-08-026-758-23 Sequence 23, Appl     |
| 18         | 41    | 45.6    | 552          | 4     | US-09-252-991A-17134 Sequence 17134, A |
| 19         | 40    | 44.4    | 95           | 4     | US-09-621-976-5302 Sequence 5302, Ap   |
| 20         | 40    | 44.4    | 263          | 4     | US-09-248-796A-20702 Sequence 20702, A |

|    |      |      |      |   |                      |                    |
|----|------|------|------|---|----------------------|--------------------|
| 21 | 40   | 44.4 | 327  | 4 | US-09-489-039A-13999 | Sequence 13999, A  |
| 22 | 40   | 44.4 | 582  | 4 | US-09-252-991A-24055 | Sequence 24055, A  |
| 23 | 40   | 44.4 | 695  | 4 | US-09-538-092-1152   | Sequence 1152, Ap  |
| 24 | 39   | 43.3 | 66   | 4 | US-10-101-464A-667   | Sequence 667, Appl |
| 25 | 39   | 43.3 | 294  | 3 | US-08-742-185-97     | Sequence 97, Appl  |
| 26 | 39   | 43.3 | 302  | 4 | US-09-252-991A-29564 | Sequence 29564, A  |
| 27 | 39   | 43.3 | 304  | 4 | US-09-328-352-5720   | Sequence 5720, Ap  |
| 28 | 39   | 43.3 | 350  | 4 | US-09-540-236-2058   | Sequence 2058, Ap  |
| 29 | 39   | 43.3 | 386  | 1 | US-08-319-621A-14    | Sequence 14, Appl  |
| 30 | 39   | 43.3 | 468  | 4 | US-09-270-767-60234  | Sequence 60234, A  |
| 31 | 39   | 43.3 | 494  | 4 | US-09-489-039A-9940  | Sequence 9940, Ap  |
| 32 | 39   | 43.3 | 758  | 4 | US-09-904-987-5      | Sequence 5, Appli  |
| 33 | 39   | 43.3 | 904  | 4 | US-09-248-796A-15547 | Sequence 15547, A  |
| 34 | 39   | 43.3 | 2410 | 4 | US-09-270-767-44775  | Sequence 44775, A  |
| 35 | 38.5 | 42.8 | 129  | 4 | US-09-270-767-33670  | Sequence 33670, A  |
| 36 | 38.5 | 42.8 | 129  | 4 | US-09-270-767-48887  | Sequence 48887, A  |
| 37 | 38.5 | 42.8 | 1835 | 3 | US-09-404-650-5      | Sequence 5, Appli  |
| 38 | 38.5 | 42.8 | 1835 | 4 | US-09-935-541-5      | Sequence 5, Appli  |
| 39 | 38   | 42.2 | 63   | 4 | US-09-489-039A-10317 | Sequence 10317, A  |
| 40 | 38   | 42.2 | 136  | 4 | US-09-288-143-190    | Sequence 190, App  |
| 41 | 38   | 42.2 | 311  | 4 | US-09-252-991A-16719 | Sequence 16719, A  |
| 42 | 38   | 42.2 | 320  | 3 | US-09-256-000-17     | Sequence 17, Appl  |
| 43 | 38   | 42.2 | 320  | 4 | US-10-034-015A-17    | Sequence 17, Appl  |
| 44 | 38   | 42.2 | 341  | 4 | US-09-252-991A-23444 | Sequence 23444, A  |
| 45 | 38   | 42.2 | 406  | 4 | US-09-252-991A-24561 | Sequence 24561, A  |

#### ALIGNMENTS

#### RESULT 1

US-08-943-667-11

; Sequence 11, Application US/08943667

; Patent No. 6734001

; GENERAL INFORMATION:

; APPLICANT: Alessi, Dario R

; TITLE OF INVENTION: ENZYME

; NUMBER OF SEQUENCES: 35

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Jaeckle Fleischmann & Muehl, LLP

; STREET: 39 State Street

; CITY: Rochester

; STATE: New York

; COUNTRY: USA

; ZIP: 14614-1310

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/943,667

; FILING DATE: 03-OCT-1997

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: GB 9705462.1

; FILING DATE: 17-MAR-1997



```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9712826.8
; FILING DATE: 19-JUN-1997
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GB 9717253.0
; FILING DATE: 15-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Braman, Susan J
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 87792.97R421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-262-3640
; TELEFAX: 716-262-4133
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
US-08-943-667-11

```

```

Query Match          54.4%; Score 49; DB 4; Length 20;
Best Local Similarity 88.9%; Pred. No. 0.17;
Matches: 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      3 DNLHQQTTP 11
        :|||||||
Db      11 ENLHQQTTP 19

```

# RESULT 2

US-09-016-000-4

```

; Sequence 4, Application US/09016000
; Patent No. 5962232

```

## GENERAL INFORMATION:

```

; APPLICANT: Hillman, Jennifer L.
; APPLICANT: Lal, Preeti
; APPLICANT: Bandman, Olga
; APPLICANT: Akerblom, Ingrid E.
; APPLICANT: Shah, Purvi
; APPLICANT: Corley, Neil C.
; APPLICANT: Guegler, Karl G.

```

```

; TITLE OF INVENTION: PROTEIN KINASE MOLECULES
; NUMBER OF SEQUENCES: 12

```

## CORRESPONDENCE ADDRESS:

```

; ADDRESSEE: Incyte Pharmaceuticals, Inc.
; STREET: 3174 Porter Drive
; CITY: Palo Alto
; STATE: CA
; COUNTRY: USA
; ZIP: 94304

```

```

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette

```

```

;      COMPUTER:  IBM Compatible
;      OPERATING SYSTEM:  DOS
;      SOFTWARE:  FastSEQ for Windows Version 2.0
;      CURRENT APPLICATION DATA:
;      APPLICATION NUMBER:  US/09/016,000
;      FILING DATE:  HERewith
;      CLASSIFICATION:
;      PRIOR APPLICATION DATA:
;      APPLICATION NUMBER:
;      FILING DATE:
;      ATTORNEY/AGENT INFORMATION:
;      NAME:  Billings, Lucy J
;      REGISTRATION NUMBER:  36,749
;      REFERENCE/DOCKET NUMBER:  PF-0465 US
;      TELECOMMUNICATION INFORMATION:
;      TELEPHONE:  650-855-0555
;      TELEFAX:  650-845-4166
;      TELEX:
;      INFORMATION FOR SEQ ID NO:  4:
;      SEQUENCE CHARACTERISTICS:
;      LENGTH:  556 amino acids
;      TYPE:  amino acid
;      STRANDEDNESS:  single
;      TOPOLOGY:  linear
;      IMMEDIATE SOURCE:
;      LIBRARY:  MMLR1DT01
;      CLONE:  472480
US-09-016-000-4

```

```

Query Match          54.4%;  Score 49;  DB 2;  Length 556;
Best Local Similarity  88.9%;  Pred. No. 6.1;
Matches      8;  Conservative    1;  Mismatches    0;  Indels    0;  Gaps    0;

```

```

Qy      3 DNLHQQTTP 11
      :|||||||
Db      348 ENLHQQTTP 356

```

```

RESULT 3
US-09-156-793D-2
; Sequence 2, Application US/09156793D
; Patent No. 6682920
; GENERAL INFORMATION:
; APPLICANT:  Stephens, Len
; APPLICANT:  Hawkins, Philip T.
; APPLICANT:  Stokoe, David
; TITLE OF INVENTION:  Compositions and Methods for Identifying PKB Kinase
; TITLE OF INVENTION:  Inhibitors
; FILE REFERENCE:  1030-US
; CURRENT APPLICATION NUMBER:  US/09/156,793D
; CURRENT FILING DATE:  1998-09-17
; PRIOR APPLICATION NUMBER:  60/060,190
; PRIOR FILING DATE:  1997-09-26
; NUMBER OF SEQ ID NOS:  2
; SOFTWARE:  PatentIn Ver. 2.0
; SEQ ID NO 2
; LENGTH:  556

```

; TYPE: PRT  
; ORGANISM: PKB Kinase  
US-09-156-793D-2

Query Match 54.4%; Score 49; DB 4; Length 556;  
Best Local Similarity 88.9%; Pred. No. 6.1;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTPP 11  
:|||||||  
Db 348 ENLHQQTPP 356

RESULT 4

US-08-943-667-1

; Sequence 1, Application US/08943667  
; Patent No. 6734001  
; GENERAL INFORMATION:  
; APPLICANT: Alessi, Dario R  
; TITLE OF INVENTION: ENZYME  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jaeckle Fleischmann & Mugel, LLP  
; STREET: 39 State Street  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14614-1310  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/943,667  
; FILING DATE: 03-OCT-1997  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9705462.1  
; FILING DATE: 17-MAR-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9712826.8  
; FILING DATE: 19-JUN-1997  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: GB 9717253.0  
; FILING DATE: 15-AUG-1997  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Braman, Susan J  
; REGISTRATION NUMBER: 34,103  
; REFERENCE/DOCKET NUMBER: 87792.97R421  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 716-262-3640  
; TELEFAX: 716-262-4133  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 556 amino acids  
; TYPE: amino acid

; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: Homo sapiens

US-08-943-667-1

Query Match 54.4%; Score 49; DB 4; Length 556;  
Best Local Similarity 88.9%; Pred. No. 6.1;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
:|||||||  
Db 348 ENLHQQTTP 356

RESULT 5

US-09-270-767-37903

; Sequence 37903, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 37903  
; LENGTH: 141  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-37903

Query Match 48.9%; Score 44; DB 4; Length 141;  
Best Local Similarity 57.1%; Pred. No. 9.1;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 LDNLHQQTTPDGF 15  
|||:| |: |||  
Db 40 LDNIHSQSYMDDFG 53

RESULT 6

US-09-270-767-53120

; Sequence 53120, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0

; SEQ ID NO 53120  
; LENGTH: 141  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-53120

Query Match 48.9%; Score 44; DB 4; Length 141;  
Best Local Similarity 57.1%; Pred. No. 9.1;  
Matches 8; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

Qy 2 LDNLHQQTTPPDGFG 15  
|||:| |: | |  
Db 40 LDNIHSQSYMDDFG 53

RESULT 7

US-09-270-767-56695  
; Sequence 56695, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 56695  
; LENGTH: 152  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-56695

Query Match 48.9%; Score 44; DB 4; Length 152;  
Best Local Similarity 66.7%; Pred. No. 9.8;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 NLHQQTTPPDGFG 15  
||:|||| |:| |  
Db 106 NLNQQTTPNGLG 117

RESULT 8

US-09-270-767-31989  
; Sequence 31989, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 31989  
; LENGTH: 226  
; TYPE: PRT

; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-31989

Query Match 48.9%; Score 44; DB 4; Length 226;  
Best Local Similarity 66.7%; Pred. No. 15;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 NLHQQTTPPDGFG 15  
||:|||| |:| |  
Db 8 NLNQQTMPNGLG 19

RESULT 9

US-09-270-767-47206  
; Sequence 47206, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 47206  
; LENGTH: 226  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
; FEATURE:  
; OTHER INFORMATION: Xaa means any amino acid  
US-09-270-767-47206

Query Match 48.9%; Score 44; DB 4; Length 226;  
Best Local Similarity 66.7%; Pred. No. 15;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 NLHQQTTPPDGFG 15  
||:|||| |:| |  
Db 8 NLNQQTMPNGLG 19

RESULT 10

US-09-270-767-41474  
; Sequence 41474, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:  
; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 41474  
; LENGTH: 389

; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-41474

Query Match 48.9%; Score 44; DB 4; Length 389;  
Best Local Similarity 66.7%; Pred. No. 27;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 NLHQQTTPDGFG 15  
||:|||| |:| |  
Db 343 NLNQQTMPNGLG 354

RESULT 11

US-09-252-991A-33044  
; Sequence 33044, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 33044  
; LENGTH: 390  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-33044

Query Match 48.9%; Score 44; DB 4; Length 390;  
Best Local Similarity 75.0%; Pred. No. 27;  
Matches 9; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 5 LHQQTTPDGFGGR 16  
| || |||| |  
Db 312 LWQQHPPDGQGR 323

RESULT 12

US-09-107-532A-6795  
; Sequence 6795, Application US/09107532A  
; Patent No. 6583275  
; GENERAL INFORMATION:  
; APPLICANT: Lynn A Doucette-Stamm and David Bush  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
; ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND  
THERAPEUTICS  
; NUMBER OF SEQUENCES: 7310  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

```

; STREET: 100 Beaver Street
; CITY: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
;
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD-ROM ISO9660
; COMPUTER: PC
; OPERATING SYSTEM: <Unknown>
; SOFTWARE: ASCII
;
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532A
; FILING DATE: 30-Jun-1998
;
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/085,598
; FILING DATE: 14 May 1998
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
;
; ATTORNEY/AGENT INFORMATION:
; NAME: Ariniello, Pamela Deneke
; REGISTRATION NUMBER: 40,489
; REFERENCE/DOCKET NUMBER: GTC-012
;
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (781)893-5007
; TELEFAX: (781)893-8277
;
; INFORMATION FOR SEQ ID NO: 6795:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 289 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: YES
; ORIGINAL SOURCE:
; ORGANISM: Enterococcus faecium
;
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (B) LOCATION 1...289
;
; SEQUENCE DESCRIPTION: SEQ ID NO: 6795:
US-09-107-532A-6795

```

```

Query Match          47.2%; Score 42.5; DB 4; Length 289;
Best Local Similarity 71.4%; Pred. No. 35;
Matches 10; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

```

```

Qy      2 LDNLHQQTTPDGFG 15
        | ||||| | ||
Db      112 LANLHQQTAPQ-FG 124

```

```

RESULT 13
US-09-543-681A-4194
; Sequence 4194, Application US/09543681A
; Patent No. 6605709
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS
MIRABILIS FOR

```



; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 4194  
; LENGTH: 589  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-09-543-681A-4194

Query Match 46.7%; Score 42; DB 4; Length 589;  
Best Local Similarity 57.1%; Pred. No. 90;  
Matches 8; Conservative 1; Mismatches 5; Indels 0; Gaps 0;

Qy 1 LLDNLHQQTTPDGF 14  
|:| | | | | |  
Db 55 LIDTLDQQIPTDSF 68

RESULT 14

US-09-543-681A-7058  
; Sequence 7058, Application US/09543681A  
; Patent No. 6605709  
; GENERAL INFORMATION:  
; APPLICANT: GARY BRETON  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS  
MIRABILIS FOR  
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1002-001  
; CURRENT APPLICATION NUMBER: US/09/543,681A  
; CURRENT FILING DATE: 2000-04-05  
; PRIOR APPLICATION NUMBER: US 60/128,706  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 8344  
; SEQ ID NO 7058  
; LENGTH: 1092  
; TYPE: PRT  
; ORGANISM: Proteus mirabilis  
US-C9-543-681A-7058

Query Match 46.7%; Score 42; DB 4; Length 1092;  
Best Local Similarity 58.3%; Pred. No. 1.8e+02;  
Matches 7; Conservative 3; Mismatches 2; Indels 0; Gaps 0;

Qy 1 LLDNLHQQTTPD 12  
| |:::| | | |  
Db 141 LYDDIYQGTPPD 152

RESULT 15

US-09-270-767-61771  
; Sequence 61771, Application US/09270767  
; Patent No. 6703491  
; GENERAL INFORMATION:

; APPLICANT: Homburger et al.  
; TITLE OF INVENTION: Nucleic acids and proteins of Drosophila melanogaster  
; FILE REFERENCE: File Reference: 7326-094  
; CURRENT APPLICATION NUMBER: US/09/270,767  
; CURRENT FILING DATE: 1999-03-17  
; NUMBER OF SEQ ID NOS: 62517  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 61771  
; LENGTH: 31  
; TYPE: PRT  
; ORGANISM: Drosophila melanogaster  
US-09-270-767-61771

Query Match 45.6%; Score 41; DB 4; Length 31;  
Best Local Similarity 50.0%; Pred. No. 5.5;  
Matches 8; Conservative 1; Mismatches 7; Indels 0; Gaps 0;

Qy 1 LLDNLHQQTTPDGFGR 16  
| | | | : | | | |  
Db 14 LKDNFHKQTKPSAKNR 29

Search completed: January 31, 2005, 13:25:09  
Job time : 29.3636 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

CM protein - protein search, using sw model

Run on: January 31, 2005, 13:22:56 ; Search time 91.2727 Seconds  
(without alignments)  
63.334 Million cell updates/sec

Title: US-10-067-620-6  
Perfect score: 90  
Sequence: 1 LLDNLHQQTTPDGFGR 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*

4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*  
 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*  
 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*  
 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*  
 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*  
 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*  
 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*  
 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*  
 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
 14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*  
 17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
 score greater than or equal to the score of the result being printed,  
 and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result<br>No. | Score | %<br>Query |        | DB | ID                 | Description       |
|---------------|-------|------------|--------|----|--------------------|-------------------|
|               |       | Match      | Length |    |                    |                   |
| 1             | 90    | 100.0      | 16     | 14 | US-10-067-484-6    | Sequence 6, Appli |
| 2             | 90    | 100.0      | 16     | 14 | US-10-067-620-6    | Sequence 6, Appli |
| 3             | 49    | 54.4       | 20     | 14 | US-10-190-012-11   | Sequence 11, Appl |
| 4             | 49    | 54.4       | 35     | 9  | US-09-205-658-199  | Sequence 199, App |
| 5             | 49    | 54.4       | 35     | 10 | US-09-963-693-199  | Sequence 199, App |
| 6             | 49    | 54.4       | 113    | 9  | US-09-205-658-173  | Sequence 173, App |
| 7             | 49    | 54.4       | 113    | 10 | US-09-963-693-173  | Sequence 173, App |
| 8             | 49    | 54.4       | 285    | 14 | US-10-217-155A-13  | Sequence 13, Appl |
| 9             | 49    | 54.4       | 285    | 15 | US-10-217-574-13   | Sequence 13, Appl |
| 10            | 49    | 54.4       | 285    | 15 | US-10-217-555-13   | Sequence 13, Appl |
| 11            | 49    | 54.4       | 319    | 17 | US-10-473-127-1375 | Sequence 1375, Ap |
| 12            | 49    | 54.4       | 319    | 17 | US-10-473-127-1386 | Sequence 1386, Ap |
| 13            | 49    | 54.4       | 361    | 16 | US-10-664-421-106  | Sequence 106, App |
| 14            | 49    | 54.4       | 468    | 17 | US-10-473-127-1371 | Sequence 1371, Ap |
| 15            | 49    | 54.4       | 468    | 17 | US-10-473-127-1385 | Sequence 1385, Ap |
| 16            | 49    | 54.4       | 506    | 17 | US-10-473-127-1381 | Sequence 1381, Ap |
| 17            | 49    | 54.4       | 535    | 17 | US-10-473-127-1379 | Sequence 1379, Ap |
| 18            | 49    | 54.4       | 556    | 9  | US-09-771-161A-245 | Sequence 245, App |
| 19            | 49    | 54.4       | 556    | 14 | US-10-190-012-1    | Sequence 1, Appli |
| 20            | 49    | 54.4       | 556    | 14 | US-10-376-554-6    | Sequence 6, Appli |
| 21            | 49    | 54.4       | 556    | 16 | US-10-704-921-14   | Sequence 14, Appl |
| 22            | 49    | 54.4       | 556    | 17 | US-10-473-127-1372 | Sequence 1372, Ap |
| 23            | 49    | 54.4       | 556    | 17 | US-10-473-127-1373 | Sequence 1373, Ap |
| 24            | 49    | 54.4       | 556    | 17 | US-10-473-127-1374 | Sequence 1374, Ap |
| 25            | 49    | 54.4       | 556    | 17 | US-10-473-127-1377 | Sequence 1377, Ap |
| 26            | 49    | 54.4       | 556    | 17 | US-10-473-127-1378 | Sequence 1378, Ap |
| 27            | 49    | 54.4       | 556    | 17 | US-10-473-127-1380 | Sequence 1380, Ap |
| 28            | 49    | 54.4       | 556    | 17 | US-10-473-127-1382 | Sequence 1382, Ap |
| 29            | 49    | 54.4       | 556    | 17 | US-10-473-127-1383 | Sequence 1383, Ap |
| 30            | 49    | 54.4       | 556    | 17 | US-10-473-127-1384 | Sequence 1384, Ap |

|    |    |      |     |    |                      |                   |
|----|----|------|-----|----|----------------------|-------------------|
| 31 | 49 | 54.4 | 556 | 17 | US-10-723-860-2053   | Sequence 2053, Ap |
| 32 | 47 | 52.2 | 114 | 15 | US-10-424-599-221678 | Sequence 221678,  |
| 33 | 46 | 51.1 | 63  | 15 | US-10-424-599-233381 | Sequence 233381,  |
| 34 | 45 | 50.0 | 58  | 17 | US-10-425-115-209997 | Sequence 209997,  |
| 35 | 45 | 50.0 | 591 | 17 | US-10-425-115-331152 | Sequence 331152,  |
| 36 | 45 | 50.0 | 616 | 15 | US-10-425-114-72559  | Sequence 72559, A |
| 37 | 45 | 50.0 | 652 | 16 | US-10-437-963-195442 | Sequence 195442,  |
| 38 | 44 | 48.9 | 82  | 14 | US-10-106-698-8178   | Sequence 8178, Ap |
| 39 | 44 | 48.9 | 170 | 16 | US-10-437-963-122298 | Sequence 122298,  |
| 40 | 44 | 48.9 | 591 | 15 | US-10-282-122A-69558 | Sequence 69558, A |
| 41 | 43 | 47.8 | 50  | 16 | US-10-437-963-106215 | Sequence 106215,  |
| 42 | 43 | 47.8 | 140 | 15 | US-10-424-599-185380 | Sequence 185380,  |
| 43 | 43 | 47.8 | 242 | 16 | US-10-437-963-138422 | Sequence 138422,  |
| 44 | 43 | 47.8 | 472 | 17 | US-10-739-930-9412   | Sequence 9412, Ap |
| 45 | 43 | 47.8 | 596 | 17 | US-10-739-930-11023  | Sequence 11023, A |

#### ALIGNMENTS

##### RESULT 1

US-10-067-484-6

```
; Sequence 6, Application US/10067484
; Publication No. US20030170763A1
; GENERAL INFORMATION:
; APPLICANT: Buchanan, Bob B.
; APPLICANT: del Val, Gregorio
; APPLICANT: Frick, Oscar L.
; TITLE OF INVENTION: RAGWEED ALLERGENS
; FILE REFERENCE: 416272000200
; CURRENT APPLICATION NUMBER: US/10/067,484
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 60/266,686
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 16
; TYPE: PRT
; ORGANISM: Ragweed
US-10-067-484-6
```

```
Query Match          100.0%; Score 90; DB 14; Length 16;
Best Local Similarity 100.0%; Pred. No. 2e-07;
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1 LLDNLHQQTTPPDGFGR 16
         |||||
Db      1 LLDNLHQQTTPPDGFGR 16
```

##### RESULT 2

US-10-067-620-6

```
; Sequence 6, Application US/10067620
; Publication No. US20030180225A1
; GENERAL INFORMATION:
; APPLICANT: Buchanan, Bob B.
```

; APPLICANT: del Val, Gregorio  
; APPLICANT: Frick, Oscar L.  
; APPLICANT: Teuber, Suzanne S.  
; TITLE OF INVENTION: WALNUT AND RYEGRASS ALLERGENS  
; FILE REFERENCE: 416272003400  
; CURRENT APPLICATION NUMBER: US/10/067,620  
; CURRENT FILING DATE: 2002-02-04  
; PRIOR APPLICATION NUMBER: US 60/266,686  
; PRIOR FILING DATE: 2001-02-05  
; NUMBER OF SEQ ID NOS: 11  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 16  
; TYPE: PRT  
; ORGANISM: Ragweed  
US-10-067-620-6

Query Match 100.0%; Score 90; DB 14; Length 16;  
Best Local Similarity 100.0%; Pred. No. 2e-07;  
Matches 16; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LLDNLHQQTTPPDGFR 16  
| | | | | | | | | | | | | | | |  
Db 1 LLDNLHQQTTPPDGFR 16

RESULT 3

US-10-190-012-11

; Sequence 11, Application US/10190012  
; Publication No. US20030108971A1

; GENERAL INFORMATION:

; APPLICANT: Alessi, Dario R  
; TITLE OF INVENTION: ENZYME  
; NUMBER OF SEQUENCES: 35  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Jaeckle Fleischmann & Mugel, LLP  
; STREET: 39 State Street  
; CITY: Rochester  
; STATE: New York  
; COUNTRY: USA  
; ZIP: 14614-1310

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/190,012  
; FILING DATE: 05-Jul-2002  
; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US/08/943,667  
; FILING DATE: 03-OCT-1997  
; APPLICATION NUMBER: GB 9705462.1  
; FILING DATE: 17-MAR-1997  
; APPLICATION NUMBER: GB 9712826.8  
; FILING DATE: 19-JUN-1997

```

; APPLICATION NUMBER: GB 9717253.0
; FILING DATE: 15-AUG-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Braman, Susan J
; REGISTRATION NUMBER: 34,103
; REFERENCE/DOCKET NUMBER: 87792.97R421
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 716-262-3640
; TELEFAX: 716-262-4133
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 20 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: internal
; SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-190-012-11

```

```

Query Match          54.4%; Score 49; DB 14; Length 20;
Best Local Similarity 88.9%; Pred. No. 0.91;
Matches      8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      3 DNLHQQTPP 11
        :|||||||
Db      11 ENLHQQTPP 19

```

#### RESULT 4

```

US-09-205-658-199
; Sequence 199, Application US/09205658
; Patent No. US20010029617A1
; GENERAL INFORMATION:
; APPLICANT: Ruvkun, Gary
; APPLICANT: Ogg, Scott
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS
; FILE REFERENCE: 00786/351004
; CURRENT APPLICATION NUMBER: US/09/205,658
; CURRENT FILING DATE: 1998-12-03
; EARLIER APPLICATION NUMBER: 08/857,076
; EARLIER FILING DATE: 1997-05-15
; EARLIER APPLICATION NUMBER: 08/888,534
; EARLIER FILING DATE: 1997-07-07
; EARLIER APPLICATION NUMBER: US98/10080
; EARLIER FILING DATE: 1998-05-15
; NUMBER OF SEQ ID NOS: 328
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 199
; LENGTH: 35
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-205-658-199

```

Query Match 54.4%; Score 49; DB 9; Length 35;  
Best Local Similarity 88.9%; Pred. No. 1.7;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
:|||||||  
Db 13 ENLHQQTTP 21

RESULT 5

US-09-963-693-199

; Sequence 199, Application US/09963693  
; Publication No. US20030181364A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruvkun, Gary  
; APPLICANT: Ogg, Scott  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
; FILE REFERENCE: 00786/351004  
; CURRENT APPLICATION NUMBER: US/09/963,693  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/205,658  
; PRIOR FILING DATE: 1998-12-03  
; PRIOR APPLICATION NUMBER: 08/857,076  
; PRIOR FILING DATE: 1997-05-15  
; PRIOR APPLICATION NUMBER: 08/888,534  
; PRIOR FILING DATE: 1997-07-07  
; PRIOR APPLICATION NUMBER: US98/10080  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 328  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 199  
; LENGTH: 35  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-963-693-199

Query Match 54.4%; Score 49; DB 10; Length 35;  
Best Local Similarity 88.9%; Pred. No. 1.7;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
:|||||||  
Db 13 ENLHQQTTP 21

RESULT 6

US-09-205-658-173

; Sequence 173, Application US/09205658  
; Patent No. US20010029617A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruvkun, Gary  
; APPLICANT: Ogg, Scott  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
; FILE REFERENCE: 00786/351004  
; CURRENT APPLICATION NUMBER: US/09/205,658

; CURRENT FILING DATE: 1998-12-03  
; EARLIER APPLICATION NUMBER: 08/857,076  
; EARLIER FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: 08/888,534  
; EARLIER FILING DATE: 1997-07-07  
; EARLIER APPLICATION NUMBER: US98/10080  
; EARLIER FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 328  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 173  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Mus musculus or Homo sapiens  
US-09-205-658-173

Query Match 54.4%; Score 49; DB 9; Length 113;  
Best Local Similarity 88.9%; Pred. No. 5.8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
:|||||||  
Db 86 ENLHQQTTP 94

RESULT 7

US-09-963-693-173  
; Sequence 173, Application US/09963693  
; Publication No. US20030181364A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruvkun, Gary  
; APPLICANT: Ogg, Scott  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
; FILE REFERENCE: 00786/351004  
; CURRENT APPLICATION NUMBER: US/09/963,693  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/205,658  
; PRIOR FILING DATE: 1998-12-03  
; PRIOR APPLICATION NUMBER: 08/857,076  
; PRIOR FILING DATE: 1997-05-15  
; PRIOR APPLICATION NUMBER: 08/888,534  
; PRIOR FILING DATE: 1997-07-07  
; PRIOR APPLICATION NUMBER: US98/10080  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 328  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 173  
; LENGTH: 113  
; TYPE: PRT  
; ORGANISM: Mus musculus or Homo sapiens  
US-09-963-693-173

Query Match 54.4%; Score 49; DB 10; Length 113;  
Best Local Similarity 88.9%; Pred. No. 5.8;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11



Db                   :|||||||  
86 ENLHQQTTP 94

RESULT 8

US-10-217-155A-13  
; Sequence 13, Application US/10217155A  
; Publication No. US20030065855A1  
; GENERAL INFORMATION:  
; APPLICANT: Barford, David  
; APPLICANT: Yang, Jing  
; APPLICANT: Hemmings, Brian A  
; APPLICANT: Cron, Peter D  
; TITLE OF INVENTION: Kinase Crystal Structures and Materials and Methods for  
; TITLE OF INVENTION: Kinase Activation  
; FILE REFERENCE: 44236  
; CURRENT APPLICATION NUMBER: US/10/217,155A  
; CURRENT FILING DATE: 2002-08-14  
; PRIOR APPLICATION NUMBER: GB 0119860.5  
; PRIOR FILING DATE: 2001-08-14  
; PRIOR APPLICATION NUMBER: GB 0209985.1  
; PRIOR FILING DATE: 2002-05-01  
; NUMBER OF SEQ ID NOS: 40  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 13  
; LENGTH: 285  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Sequence source  
; OTHER INFORMATION: uncertain  
US-10-217-155A-13

Query Match                   54.4%; Score 49; DB 14; Length 285;  
Best Local Similarity       88.9%; Pred. No. 16;  
Matches       8; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy                   3 DNLHQQTTP 11  
                     :|||||||  
Db                   273 ENLHQQTTP 281

RESULT 9

US-10-217-574-13  
; Sequence 13, Application US/10217574  
; Publication No. US20040005687A1  
; GENERAL INFORMATION:  
; APPLICANT: Barford, David  
; APPLICANT: Yang, Jing  
; APPLICANT: Hemmings, Brian A  
; APPLICANT: Cron, Peter D  
; TITLE OF INVENTION: Kinase Crystal Structures  
; FILE REFERENCE: 44237  
; CURRENT APPLICATION NUMBER: US/10/217,574  
; CURRENT FILING DATE: 2002-12-23  
; PRIOR APPLICATION NUMBER: GB 0119860.5  
; PRIOR FILING DATE: 2001-08-14

; PRIOR APPLICATION NUMBER: GB 0209985.1  
 ; PRIOR FILING DATE: 2002-05-01  
 ; PRIOR APPLICATION NUMBER: GB 0216215.4  
 ; PRIOR FILING DATE: 2002-07-12  
 ; NUMBER OF SEQ ID NOS: 46  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 13  
 ; LENGTH: 285  
 ; TYPE: PRT  
 ; ORGANISM: Unknown Organism  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Unknown Organism: Sequence source  
 ; OTHER INFORMATION: uncertain  
 US-10-217-574-13

Query Match 54.4%; Score 49; DB 15; Length 285;  
 Best Local Similarity 88.9%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTPP 11  
 :|||||||  
 Db 273 ENLHQQTPP 281

# RESULT 10

US-10-217-555-13  
 ; Sequence 13, Application US/10217555  
 ; Publication No. US20040009569A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Barford, David  
 ; APPLICANT: Yang, Jing  
 ; APPLICANT: Hemmings, Brian A  
 ; APPLICANT: Cron, Peter D  
 ; TITLE OF INVENTION: Kinase Crystal Structures and Materials and Methods for  
 ; TITLE OF INVENTION: Kinase Activation  
 ; FILE REFERENCE: 44236  
 ; CURRENT APPLICATION NUMBER: US/10/217,555  
 ; CURRENT FILING DATE: 2002-12-05  
 ; PRIOR APPLICATION NUMBER: GB 0119860.5  
 ; PRIOR FILING DATE: 2001-08-14  
 ; PRIOR APPLICATION NUMBER: GB 0209985.1  
 ; PRIOR FILING DATE: 2002-05-01  
 ; NUMBER OF SEQ ID NOS: 40  
 ; SOFTWARE: PatentIn Ver. 2.1  
 ; SEQ ID NO 13  
 ; LENGTH: 285  
 ; TYPE: PRT  
 ; ORGANISM: Unknown Organism  
 ; FEATURE:  
 ; OTHER INFORMATION: Description of Unknown Organism: Sequence source  
 ; OTHER INFORMATION: uncertain  
 US-10-217-555-13

Query Match 54.4%; Score 49; DB 15; Length 285;  
 Best Local Similarity 88.9%; Pred. No. 16;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTPP 11  
:|||||||  
Db 273 ENLHQQTPP 281

RESULT 11

US-10-473-127-1375  
; Sequence 1375, Application US/10473127  
; Publication No. US20040236091A1  
; GENERAL INFORMATION:  
; APPLICANT: Zycos Inc.  
; TITLE OF INVENTION: TRANSLATIONAL PROFILING  
; FILE REFERENCE: 08191-026W01  
; CURRENT APPLICATION NUMBER: US/10/473,127  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: 60/279,495  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: 60/292,544  
; PRIOR FILING DATE: 2001-05-21  
; PRIOR APPLICATION NUMBER: 60/310,801  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: 60/326,370  
; PRIOR FILING DATE: 2001-10-01  
; PRIOR APPLICATION NUMBER: 60/336,780  
; PRIOR FILING DATE: 2001-12-04  
; PRIOR APPLICATION NUMBER: 60/358,985  
; PRIOR FILING DATE: 2002-02-20  
; NUMBER OF SEQ ID NOS: 2041  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1375  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-473-127-1375

Query Match 54.4%; Score 49; DB 17; Length 319;  
Best Local Similarity 88.9%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTPP 11  
:|||||||  
Db 111 ENLHQQTPP 119

RESULT 12

US-10-473-127-1386  
; Sequence 1386, Application US/10473127  
; Publication No. US20040236091A1  
; GENERAL INFORMATION:  
; APPLICANT: Zycos Inc.  
; TITLE OF INVENTION: TRANSLATIONAL PROFILING  
; FILE REFERENCE: 08191-026W01  
; CURRENT APPLICATION NUMBER: US/10/473,127  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: 60/279,495  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: 60/292,544

; PRIOR FILING DATE: 2001-05-21  
; PRIOR APPLICATION NUMBER: 60/310,801  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: 60/326,370  
; PRIOR FILING DATE: 2001-10-01  
; PRIOR APPLICATION NUMBER: 60/336,780  
; PRIOR FILING DATE: 2001-12-04  
; PRIOR APPLICATION NUMBER: 60/358,985  
; PRIOR FILING DATE: 2002-02-20  
; NUMBER OF SEQ ID NOS: 2041  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1386  
; LENGTH: 319  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-473-127-1386

Query Match 54.4%; Score 49; DB 17; Length 319;  
Best Local Similarity 88.9%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
:|||||||  
Db 111 ENLHQQTTP 119

RESULT 13

US-10-664-421-106

; Sequence 106, Application US/10654421  
; Publication No. US20040142864A1  
; GENERAL INFORMATION:  
; APPLICANT: BREMER, RYAN  
; APPLICANT: IBRAHIM, PRABHA  
; APPLICANT: KUMAR, ABHINAV  
; APPLICANT: MANDIYAN, VALSAN  
; APPLICANT: MILBURN, MICHAEL V.  
; TITLE OF INVENTION: CRYSTAL STRUCTURE OF PIM-1 KINASE  
; FILE REFERENCE: 039363/0703  
; CURRENT APPLICATION NUMBER: US/10/664,421  
; CURRENT FILING DATE: 2003-09-16  
; PRIOR APPLICATION NUMBER: 60/412,341  
; PRIOR FILING DATE: 2002-09-20  
; PRIOR APPLICATION NUMBER: 60/411,398  
; PRIOR FILING DATE: 2002-09-16  
; NUMBER OF SEQ ID NOS: 169  
; SOFTWARE: PatentIn Ver. 3.2  
; SEQ ID NO 106  
; LENGTH: 361  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-664-421-106

Query Match 54.4%; Score 49; DB 16; Length 361;  
Best Local Similarity 88.9%; Pred. No. 20;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11

Db                   :|||||||  
317 ENLHQQTTP 325

RESULT 14

US-10-473-127-1371  
; Sequence 1371, Application US/10473127  
; Publication No. US20040236091A1  
; GENERAL INFORMATION:  
; APPLICANT: Zycos Inc.  
; TITLE OF INVENTION: TRANSLATIONAL PROFILING  
; FILE REFERENCE: 08191-026W01  
; CURRENT APPLICATION NUMBER: US/10/473,127  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: 60/279,495  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: 60/292,544  
; PRIOR FILING DATE: 2001-05-21  
; PRIOR APPLICATION NUMBER: 60/310,801  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: 60/326,370  
; PRIOR FILING DATE: 2001-10-01  
; PRIOR APPLICATION NUMBER: 60/336,780  
; PRIOR FILING DATE: 2001-12-04  
; PRIOR APPLICATION NUMBER: 60/358,985  
; PRIOR FILING DATE: 2002-02-20  
; NUMBER OF SEQ ID NOS: 2041  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1371  
; LENGTH: 468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-473-127-1371

Query Match                   54.4%; Score 49; DB 17; Length 468;  
Best Local Similarity       88.9%; Pred. No. 26;  
Matches       8; Conservative   1; Mismatches   0; Indels   0; Gaps   0;

Qy                   3 DNLHQQTTP 11  
                     :|||||||  
Db                   260 ENLHQQTTP 268

RESULT 15

US-10-473-127-1385  
; Sequence 1385, Application US/10473127  
; Publication No. US20040236091A1  
; GENERAL INFORMATION:  
; APPLICANT: Zycos Inc.  
; TITLE OF INVENTION: TRANSLATIONAL PROFILING  
; FILE REFERENCE: 08191-026W01  
; CURRENT APPLICATION NUMBER: US/10/473,127  
; CURRENT FILING DATE: 2003-09-26  
; PRIOR APPLICATION NUMBER: 60/279,495  
; PRIOR FILING DATE: 2001-03-28  
; PRIOR APPLICATION NUMBER: 60/292,544  
; PRIOR FILING DATE: 2001-05-21

; PRIOR APPLICATION NUMBER: 60/310,801  
; PRIOR FILING DATE: 2001-08-08  
; PRIOR APPLICATION NUMBER: 60/326,370  
; PRIOR FILING DATE: 2001-10-01  
; PRIOR APPLICATION NUMBER: 60/336,780  
; PRIOR FILING DATE: 2001-12-04  
; PRIOR APPLICATION NUMBER: 60/358,985  
; PRIOR FILING DATE: 2002-02-20  
; NUMBER OF SEQ ID NOS: 2041  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 1385  
; LENGTH: 468  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-10-473-127-1385

Query Match 54.4%; Score 49; DB 17; Length 468;  
Best Local Similarity 88.9%; Pred. No. 26;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
:|||||||  
Db 260 ENLHQQTTP 268

Search completed: January 31, 2005, 13:44:50  
Job time : 92.2727 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 13:07:55 ; Search time 21.0909 Seconds  
(without alignments)  
72.992 Million cell updates/sec

Title: US-10-067-620-6  
Perfect score: 90  
Sequence: 1 LLDNLHQQTTPDGFR 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*

1: pir1:\*  
 2: pir2:\*  
 3: pir3:\*  
 4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result<br>No. | Score | %              |        | DB | ID     | Description        |
|---------------|-------|----------------|--------|----|--------|--------------------|
|               |       | Query<br>Match | Length |    |        |                    |
| 1             | 46    | 51.1           | 2027   | 2  | S60123 | hypothetical prote |
| 2             | 46    | 51.1           | 2056   | 2  | G88564 | protein R10E11.1 [ |
| 3             | 43    | 47.8           | 445    | 2  | T01214 | hypothetical prote |
| 4             | 42    | 46.7           | 310    | 2  | T01090 | hypothetical prote |
| 5             | 42    | 46.7           | 562    | 2  | T15131 | hypothetical prote |
| 6             | 42    | 46.7           | 911    | 2  | S28498 | gene p120 protein  |
| 7             | 41    | 45.6           | 182    | 1  | IVMSB  | interferon beta pr |
| 8             | 41    | 45.6           | 184    | 2  | JC5424 | interferon beta pr |
| 9             | 41    | 45.6           | 357    | 2  | JQ2174 | hypothetical 39.2K |
| 10            | 41    | 45.6           | 357    | 2  | S27909 | hypothetical prote |
| 11            | 41    | 45.6           | 520    | 1  | S50576 | probable aldehyde  |
| 12            | 40    | 44.4           | 177    | 2  | E85741 | hypothetical prote |
| 13            | 40    | 44.4           | 311    | 2  | D85631 | hypothetical prote |
| 14            | 40    | 44.4           | 349    | 2  | T07396 | probable outward r |
| 15            | 40    | 44.4           | 387    | 2  | JN0793 | adaptive-response  |
| 16            | 40    | 44.4           | 388    | 2  | H85692 | hypothetical prote |
| 17            | 40    | 44.4           | 391    | 2  | C90908 | probable host spec |
| 18            | 40    | 44.4           | 391    | 2  | H90968 | probable host spec |
| 19            | 40    | 44.4           | 391    | 2  | H90996 | probable host spec |
| 20            | 40    | 44.4           | 401    | 2  | G87552 | conserved hypothet |
| 21            | 40    | 44.4           | 405    | 2  | T27971 | hypothetical prote |
| 22            | 40    | 44.4           | 412    | 2  | T24023 | hypothetical prote |
| 23            | 40    | 44.4           | 416    | 2  | C95973 | probable exported  |
| 24            | 40    | 44.4           | 418    | 2  | C81712 | type III secreted  |
| 25            | 40    | 44.4           | 475    | 2  | D64799 | ybeS protein - Esc |
| 26            | 40    | 44.4           | 475    | 2  | A85565 | probable enzyme of |
| 27            | 40    | 44.4           | 475    | 2  | D90714 | probable enzyme of |
| 28            | 40    | 44.4           | 505    | 2  | T18989 | cytochrome P450 ho |
| 29            | 40    | 44.4           | 551    | 2  | AI2574 | transposase alr902 |
| 30            | 40    | 44.4           | 682    | 1  | S22700 | amphiphysin - chic |
| 31            | 40    | 44.4           | 695    | 2  | S62400 | amphiphysin (clone |
| 32            | 40    | 44.4           | 742    | 2  | T43520 | condensin complex  |
| 33            | 40    | 44.4           | 1132   | 1  | QSBPL  | host specificity p |
| 34            | 40    | 44.4           | 1132   | 2  | H90834 | host specificity p |
| 35            | 40    | 44.4           | 1137   | 2  | B90734 | probable host spec |
| 36            | 40    | 44.4           | 1138   | 2  | D85584 | probable tail comp |
| 37            | 40    | 44.4           | 1157   | 2  | A90769 | probable host spec |
| 38            | 40    | 44.4           | 1158   | 2  | F90854 | probable host spec |
| 39            | 40    | 44.4           | 1158   | 2  | G85718 | probable tail comp |
| 40            | 40    | 44.4           | 1159   | 2  | A90899 | probable host spec |
| 41            | 40    | 44.4           | 1159   | 2  | G85816 | hypothetical prote |
| 42            | 40    | 44.4           | 1165   | 2  | D85842 | probable tail fibe |
| 43            | 40    | 44.4           | 1165   | 2  | F90877 | probable host spec |

|    |    |      |      |   |        |                    |
|----|----|------|------|---|--------|--------------------|
| 44 | 40 | 44.4 | 1171 | 1 | QQKBFP | pyruvate (flavodox |
| 45 | 40 | 44.4 | 1246 | 2 | G89287 | protein H39E23.1 [ |

# ALIGNMENTS

## RESULT 1

S60123

hypothetical protein R10E11.1 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 13-Jan-1996 #sequence\_revision 01-Mar-1996 #text\_change 02-Aug-2002

C;Accession: S60123; S40713

R;Ainscough, R.; Mortimore, B.

submitted to the EMBL Data Library, November 1995

A;Reference number: S60123

A;Accession: S60123

A;Molecule type: DNA

A;Residues: 1-2027 <AIN>

A;Cross-references: EMBL:Z29095; NID:g436453; PID:g1067032

A;Note: this is a revision to the sequence from reference S40713

R;Ainscough, R.; Mortimore, B.

submitted to the EMBL Data Library, December 1993

A;Reference number: S40713

A;Accession: S40713

A;Molecule type: DNA

A;Residues: 1-466, 'CKYITRRVASFSLSGK', 467, 'FEHFR', 474-475, 'KRLFPPKISLHSSHF', 479-1986, 'GQ' <AIW>

A;Cross-references: EMBL:Z29095

A;Note: this sequence has been revised in reference S60123

C;Genetics:

A;Introns: 14/1; 39/3; 302/3; 424/3; 467/1; 517/1; 688/1; 1759/1; 1828/2; 1892/3; 1964/3; 1987/1

C;Superfamily: transcription coactivator CREB-binding protein; bromodomain homology

F;889-946/Domain: bromodomain homology <BRO>

|                       |        |               |       |                                 |
|-----------------------|--------|---------------|-------|---------------------------------|
| Query Match           | 51.1%; | Score 46;     | DB 2; | Length 2027;                    |
| Best Local Similarity | 70.0%; | Pred. No. 41; |       |                                 |
| Matches               | 7;     | Conservative  | 2;    | Mismatches 1; Indels 0; Gaps 0; |

Qy 4 NLHQQTTPPDG 13

|:| | | | | |

Db 735 NMHQQIPPNG 744

## RESULT 2

G88564

protein R10E11.1 [imported] - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 09-Jul-2004

C;Accession: G88564

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology.

A;Reference number: A75000; MUID:99069613; PMID:9851916



A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and  
 www\_sanger.ac.uk/Projects/C\_elegans/ for a list of authors  
 A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103,  
 1999; and Science 285, 1493, 1999  
 A;Accession: G88564  
 A;Status: preliminary  
 A;Molecule type: DNA  
 A;Residues: 1-2056 <STO>  
 A;Cross-references: UNIPROT:P34545; GB:chr\_III; PIDN:CAA82353.1; PID:g3979836;  
 GSPDB:GN00021; CESP:R10E11.1  
 C;Genetics:  
 A;Gene: R10E11.1  
 A;Map position: 3  
 C;Superfamily: transcription coactivator CREB-binding protein; bromodomain  
 homology

Query Match 51.1%; Score 46; DB 2; Length 2056;  
 Best Local Similarity 70.0%; Pred. No. 41;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 NLHQQTTPDG 13  
 | : | | | | : |  
 Db 735 NMHQQIPPNG 744

#### RESULT 3

T01214

hypothetical protein F6N23.21 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004

C;Accession: T01214

R;Geisel, C.

submitted to the EMBL Data Library, April 1998

A;Description: The sequence of A. thaliana F6N23.

A;Reference number: Z14281

A;Accession: T01214

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-445 <GEI>

A;Cross-references: UNIPROT:O65259; EMBL:AF058919; NID:g3047100; PID:g3047111;

GSPDB:GN00063; ATSP:F6N23.21

C;Genetics:

A;Gene: ATSP:F6N23.21

A;Map position: 5

A;Introns: 235/3; 255/2; 298/2; 343/3

Query Match 47.8%; Score 43; DB 2; Length 445;  
 Best Local Similarity 53.3%; Pred. No. 23;  
 Matches 8; Conservative 2; Mismatches 3; Indels 2; Gaps 1;

Qy 3 DNLHQQ--TPPDGFG 15  
 : | | | | | : |  
 Db 61 ENLHDPMWAPPDGYG 75

#### RESULT 4

T01090

hypothetical protein T10P11.13 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 12-Feb-1999 #sequence\_revision 12-Feb-1999 #text\_change 09-Jul-2004

C;Accession: T01090

R;Kaplan, N.; Johnson, D.; Schutz, K.; Gnoj, L.; Hoffman, J.; Till, S.; de la Bastide, M.; Granat, S.; Hameed, A.; Gottesman, T.; Hasegawa, A.; Shohdy, N.; Parnell, L.; Dedhia, N.; Johnson, A.F.; Lodhi, M.; Martienssen, R.; Chen, E.Y.; Wilson, R.; McCombie, W.R.

submitted to the EMBL Data Library, November 1998

A;Description: Sequence of A. thaliana BAC T10P11 from chromosome IV.

A;Reference number: Z14248

A;Accession: T01090

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-310 <KAP>

A;Cross-references: UNIPROT:O22768; EMBL:AC002330; NID:g2262135; PID:g3892050

A;Experimental source: cultivar Columbia

C;Genetics:

A;Map position: 4

A;Introns: 121/3; 162/3; 184/3; 206/3; 233/3

A;Note: T10P11.13

Query Match 46.7%; Score 42; DB 2; Length 310;  
Best Local Similarity 66.7%; Pred. No. 23;  
Matches 8; Conservative 0; Mismatches 4; Indels 0; Gaps 0;

QY 3 DNLHQQTTPDGF 14  
||| ||| | |  
Db 8 DNLSQDTPSDDF 19

#### RESULT 5

T15131

hypothetical protein ZC328.3 - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 09-Jul-2004

C;Accession: T15131

R;Wamsley, P.

submitted to the EMBL Data Library, April 1997

A;Description: The sequence of C. elegans cosmid ZC328.

A;Reference number: Z18298

A;Accession: T15131

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-562 <WAM>

A;Cross-references: UNIPROT:O02054; EMBL:AF000194; NID:g1946990; PID:g1946992;

PIDN:AAB52893.1; GSPDB:GN00019; CESP:ZC328.3

A;Experimental source: strain Bristol N2; clone ZC328

C;Genetics:

A;Gene: CESP:ZC328.3

A;Map position: 1

A;Introns: 37/2; 62/3; 95/1; 154/3; 179/3; 424/2; 532/2

Query Match 46.7%; Score 42; DB 2; Length 562;  
Best Local Similarity 58.3%; Pred. No. 45;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy            2 LDNLHQQTTPPDG 13  
              : |    |||||  
Db            153 IQNSQNQTTPPDG 164

RESULT 6

S28498

gene p120 protein - mouse

C;Species: Mus musculus (house mouse)

C;Date: 12-Mar-1993 #sequence\_revision 12-Mar-1993 #text\_change 09-Jul-2004

C;Accession: I48701; S28498

R;Reynolds, A.B.; Herbert, L.; Cleveland, J.L.; Berg, S.T.; Gaut, J.R.

Oncogene 7, 2439-2445, 1992

A;Title: p120, a novel substrate of protein tyrosine kinase receptors and of p60v-src, is related to cadherin-binding factors beta-catenin, plakoglobin and armadillo.

A;Reference number: I48701; MUID:93096477; PMID:1334250

A;Accession: I48701

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: mRNA

A;Residues: 1-911 <RES>

A;Cross-references: UNIPROT:P30999; EMBL:Z17804; NID:g53544; PIDN:CAA79078.1; PID:g53545

C;Genetics:

A;Gene: p120

C;Keywords: cytoskeleton

Query Match                    46.7%;    Score 42;    DB 2;    Length 911;  
Best Local Similarity    61.5%;    Pred. No. 79;  
Matches        8;    Conservative    1;    Mismatches        2;    Indels        2;    Gaps        1;

Qy            4 NLHQQTTPPDGFGFR 16  
              | |    ||||:|  
Db            205 NFHY--PPDGYGR 215

RESULT 7

IVMSB

interferon beta precursor - mouse

C;Species: Mus musculus (house mouse)

C;Date: 03-Aug-1984 #sequence\_revision 03-Aug-1984 #text\_change 09-Jul-2004

C;Accession: S02020; S04201; A01839

R;Vodjdani, G.; Coulombel, C.; Doly, J.

J. Mol. Biol. 204, 221-231, 1988

A;Title: Structure and characterization of a murine chromosomal fragment containing the interferon beta gene.

A;Reference number: S02020; MUID:89125582; PMID:3221389

A;Accession: S02020

A;Molecule type: DNA

A;Residues: 1-182 <VOD>

A;Cross-references: UNIPROT:P01575; EMBL:X14029; NID:g51550; PIDN:CAA32190.1; PID:g51551

R;Kuga, T.; Fujita, T.; Taniguchi, T.

Nucleic Acids Res. 17, 3291, 1989

A;Title: Nucleotide sequence of the mouse interferon-beta gene.

A;Reference number: S04201; MUID:89263735; PMID:2726460

A;Accession: S04201

A;Status: translation not shown  
A;Molecule type: DNA  
A;Residues: 1-182 <KUG>  
A;Cross-references: EMBL:X14455; NID:g51538; PIDN:CAA32625.1; PID:g51539  
R;Higashi, Y.; Sokawa, Y.; Watanabe, Y.; Kawade, Y.; Ohno, S.; Takaoka, C.;  
Taniguchi, T.  
J. Biol. Chem. 258, 9522-9529, 1983  
A;Title: Structure and expression of a cloned cDNA for mouse interferon-beta.  
A;Reference number: A01839; MUID:83265757; PMID:6688252  
A;Accession: A01839  
A;Molecule type: mRNA  
A;Residues: 1-182 <HIG>  
A;Cross-references: GB:K00020; NID:g194113; PIDN:AAA37891.1; PID:g309327  
C;Genetics:  
A;Map position: 4  
C;Superfamily: interferon alpha  
C;Keywords: glycoprotein  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-182/Product: interferon beta #status predicted <MAT>  
F;50,90,97/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 45.6%; Score 41; DB 1; Length 182;  
Best Local Similarity 88.9%; Pred. No. 18;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 LLDNLHQQT 9  
||| ||||  
Db 104 LLDELHQQT 112

# RESULT 8

JC5424  
interferon beta precursor - rat  
C;Species: Rattus norvegicus (Norway rat)  
C;Date: 10-Jun-1997 #sequence\_revision 18-Jul-1997 #text\_change 09-Jul-2004  
C;Accession: JC5424  
R;Yokoyama, S.; Ohishi, N.; Shamoto, M.; Watanabe, Y.; Yagi, K.  
Biochem. Biophys. Res. Commun. 232, 698-701, 1997  
A;Title: Isolation and expression of rat interferon beta gene and growth-  
inhibitory effect of its expression on rat glioma cells.  
A;Reference number: JC5424; MUID:97271387; PMID:9126338  
A;Accession: JC5424  
A;Molecule type: DNA  
A;Residues: 1-184 <YOK>  
A;Cross-references: UNIPROT:P70499; DDBJ:D87919; NID:g1616938; PIDN:BAA13502.1;  
PID:g1616939  
C;Comment: This protein exhibits characteristic antiviral and antitumor  
activities.  
C;Genetics:  
A;Gene: IFNbeta  
C;Superfamily: interferon alpha  
F;1-21/Domain: signal sequence #status predicted <SIG>  
F;22-184/Product: interferon beta #status predicted <MAT>

Query Match 45.6%; Score 41; DB 2; Length 184;  
Best Local Similarity 88.9%; Pred. No. 19;  
Matches 8; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Qy           1 LLDNLHQQT 9  
          ||| |||||  
Db           106 LLDELHQQT 114

RESULT 9

JQ2174

hypothetical 39.2K protein (clone GV-A) - garlic virus B

C;Species: garlic virus B

C;Date: 30-Sep-1993 #sequence\_revision 20-Aug-1994 #text\_change 07-May-1999

C;Accession: JQ2174

R;Sumi, S.; Tsuneyoshi, T.; Furutani, H.

J. Gen. Virol. 74, 1879-1885, 1993

A;Title: Novel rod-shaped viruses isolated from garlic, *Allium sativum*,  
possessing a unique genome organization.

A;Reference number: JQ2171; MUID:93389442; PMID:8376963

A;Accession: JQ2174

A;Molecule type: mRNA

A;Residues: 1-357 <SUM>

C;Superfamily: garlic virus B conserved hypothetical 39.2K protein

Query Match                   45.6%;   Score 41;   DB 2;   Length 357;  
Best Local Similarity   50.0%;   Pred. No. 40;  
Matches       8;   Conservative    2;   Mismatches    6;   Indels       0;   Gaps       0;

Qy           1 LLDNLHQQTTPPDGFGR 16  
          ||| :| : | | ||  
Db           271 LLDGVHSKIPMDIIGR 286

RESULT 10

S27909

hypothetical protein III - garlic virus A

C;Species: garlic virus A

C;Date: 06-Jan-1995 #sequence\_revision 06-Jan-1995 #text\_change 09-Jul-2004

C;Accession: S27909

R;Sumi, S.I.

submitted to the EMBL Data Library, July 1992

A;Reference number: S27908

A;Accession: S27909

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-357 <SUM>

A;Cross-references: UNIPROT:Q67692; EMBL:D11157

C;Superfamily: garlic virus B conserved hypothetical 39.2K protein

Query Match                   45.6%;   Score 41;   DB 2;   Length 357;  
Best Local Similarity   50.0%;   Pred. No. 40;  
Matches       8;   Conservative    2;   Mismatches    6;   Indels       0;   Gaps       0;

Qy           1 LLDNLHQQTTPPDGFGR 16  
          ||| :| : | | ||  
Db           271 LLDGVHSKIPMDIIGR 286

RESULT 11

S50576

probable aldehyde dehydrogenase (NAD) (EC 1.2.1.3) YER073w - yeast  
(*Saccharomyces cerevisiae*)

C;Species: *Saccharomyces cerevisiae*

C;Date: 28-May-1993 #sequence\_revision 31-Jan-1997 #text\_change 09-Jul-2004

C;Accession: S50576

R;Dietrich, F.S.

submitted to the EMBL Data Library, December 1994

A;Description: The sequence of *S. cerevisiae* lambda clone 3612 and cosmid 9747.

A;Reference number: S50438

A;Accession: S50576

A;Molecule type: DNA

A;Residues: 1-520 <DIE>

A;Cross-references: UNIPROT:P40047; EMBL:U18814; NID:g603309; PIDN:AAB64612.1;

PID:g603310; GSPDB:GN00005; MIPS:YER073w

A;Experimental source: strain S288C (AB972)

C;Genetics:

A;Gene: SGD:ALD5; MIPS:YER073w

A;Cross-references: SGD:S0000875

A;Map position: 5R

C;Function:

A;Description: catalyzes oxidation of an aldehyde to an acid using NAD<sup>+</sup> and water

A;Note: enzymes with this activity are involved in diverse metabolic pathways in various organisms and tissues

C;Superfamily: NAD-dependent aldehyde dehydrogenase; aldehyde dehydrogenase homology

C;Keywords: alcohol metabolism; NAD; oxidoreductase

F;82-342/Domain: aldehyde dehydrogenase homology <ALDD>

F;288,322/Active site: Glu, Cys #status predicted

Query Match 45.6%; Score 41; DB 1; Length 520;

Best Local Similarity 50.0%; Pred. No. 61;

Matches 7; Conservative 2; Mismatches 5; Indels 0; Gaps 0;

Qy 3 DNLHQQTTPDGFGR 16

:| || | |||:

Db 477 NNFHQNVPPGGFGQ 490

## RESULT 12

E85741

hypothetical protein Z2346 [imported] - *Escherichia coli* (strain O157:H7, substrain EDL933)

C;Species: *Escherichia coli*

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004

C;Accession: E85741

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic *Escherichia coli* O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: E85741

A;Status: preliminary

A;Molecule type: DNA  
A;Residues: 1-177 <STO>  
A;Cross-references: UNIPROT:Q8X426; GB:AE005174; NID:g12515336; PIDN:AAG56393.1;  
GSPDB:GN00145; UWGP:Z2346  
A;Experimental source: strain O157:H7, substrain EDL933  
C;Genetics:  
A;Gene: Z2346

Query Match 44.4%; Score 40; DB 2; Length 177;  
Best Local Similarity 75.0%; Pred. No. 26;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 7 QQTTPDGF 14  
:||||:|  
Db 74 EQTPPEGF 81

RESULT 13

D85631

hypothetical protein Z1379 [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004

C;Accession: D85631

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shac, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.; Iin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.  
Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: D85631

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-311 <STO>

A;Cross-references: UNIPROT:Q8X4D1; GB:AE005174; NID:g12514223; PIDN:AAG55512.1;  
GSPDB:GN00145; UWGP:Z1379

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: Z1379

Query Match 44.4%; Score 40; DB 2; Length 311;  
Best Local Similarity 75.0%; Pred. No. 50;  
Matches 6; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

Qy 7 QQTTPDGF 14  
:||||:|  
Db 52 EQTPPEGF 59

RESULT 14

T07396

probable outward rectifying potassium channel KCO1 - potato

C;Species: Solanum tuberosum (potato)

C;Date: 14-May-1999 #sequence\_revision 14-May-1999 #text\_change 09-Jul-2004

C;Accession: T07396

R;Czempinski, K.  
submitted to the EMBL Data Library, May 1997.  
A;Reference number: Z16007  
A;Accession: T07396  
A;Status: preliminary; translated from GB/EMBL/DDBJ  
A;Molecule type: DNA  
A;Residues: 1-349 <CZE>  
A;Cross-references: UNIPROT:O82065; EMBL:Y13048; NID:e1049403; PIDN:CAA73483.1;  
PID:e1313861  
A;Experimental source: cv. AM 80/5793; adult  
C;Genetics:  
A;Introns: 300/2  
A;Note: kcol  
C;Keywords: potassium channel; voltage-gated ion channel

Query Match 44.4%; Score 40; DB 2; Length 349;  
Best Local Similarity 56.2%; Pred. No. 58;  
Matches 9; Conservative 0; Mismatches 7; Indels 0; Gaps 0;

Qy 1 LLDNLHQQTTPDGFR 16  
||| ||| | ||  
Db 10 LLDQLHQQTQHTVGLGR 25

#### RESULT 15

JN0793  
adaptive-response sensory-kinase (EC 2.7.--.) - Synechococcus sp. (strain PCC 7942)  
N;Alternate names: signal-transduction protein  
C;Species: Synechococcus sp.  
C;Date: 24-Feb-1994 #sequence\_revision 24-Feb-1994 #text\_change 09-Jul-2004  
C;Accession: JN0793  
R;Nagaya, M.; Aiba, H.; Mizuno, T.  
Gene 131, 119-124, 1993  
A;Title: Cloning of a sensory-kinase-encoding gene that belongs to the two-component regulatory family from the cyanobacterium Synechococcus sp. PCC7942.  
A;Reference number: JN0793; MUID:93380660; PMID:8370532  
A;Accession: JN0793  
A;Molecule type: DNA  
A;Residues: 1-387 <NAG>  
A;Cross-references: UNIPROT:Q06904; DDBJ:D14056; NID:g217141; PIDN:BAA03145.1;  
PID:g217142  
C;Comment: This protein is a signal-transduction protein in adaptive-response systems in prokaryotes.  
C;Genetics:  
A;Gene: sasA  
C;Superfamily: sensory transduction system regulatory protein homolog  
C;Keywords: autophosphorylation; phosphoprotein; phosphotransferase

Query Match 44.4%; Score 40; DB 2; Length 387;  
Best Local Similarity 61.5%; Pred. No. 65;  
Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 LLDNLHQQTTPDG 13  
||| : ||| |  
Db 283 LLDNAIKYTPPGG 295



Search completed: January 31, 2005, 13:23:46  
Job time : 23.0909 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 12:56:50 ; Search time 121.818 Seconds  
(without alignments)  
75.572 Million cell updates/sec

Title: US-10-067-620-6  
Perfect score: 90  
Sequence: 1 LLDNLHQQTTPDGFR 16

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_02:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result<br>No. | Score | %<br>Query |        | DB | ID         | Description        |
|---------------|-------|------------|--------|----|------------|--------------------|
|               |       | Match      | Length |    |            |                    |
| 1             | 51    | 56.7       | 341    | 2  | Q7Q389     | Q7q389 anopheles g |
| 2             | 51    | 56.7       | 470    | 2  | Q6MHZ1     | Q6mhz1 bdellovibri |
| 3             | 51    | 56.7       | 470    | 2  | CAE78191   | Cae78191 bdellovib |
| 4             | 49    | 54.4       | 319    | 2  | Q9UPJ7     | Q9upj7 homo sapien |
| 5             | 49    | 54.4       | 367    | 2  | Q9UPJ8     | Q9upj8 homo sapien |
| 6             | 49    | 54.4       | 429    | 2  | Q9BRD5     | Q9brd5 homo sapien |
| 7             | 49    | 54.4       | 532    | 2  | Q8K3L3     | Q8k3l3 mus musculu |
| 8             | 49    | 54.4       | 551    | 2  | Q810Z4     | Q810z4 mus musculu |
| 9             | 49    | 54.4       | 556    | 1  | PDPK_HUMAN | O15530 homo sapien |
| 10            | 49    | 54.4       | 559    | 1  | PDPK_MOUSE | Q9z2a0 mus musculu |
| 11            | 49    | 54.4       | 559    | 1  | PDPK_RAT   | O55173 rattus norv |
| 12            | 46    | 51.1       | 2056   | 1  | CBP1_CAEEL | P34545 caenorhabdi |

|    |      |      |      |   |          |          |              |
|----|------|------|------|---|----------|----------|--------------|
| 13 | 45   | 50.0 | 183  | 2 | Q7U666   | Q7u666   | synechococc  |
| 14 | 45   | 50.0 | 332  | 2 | Q7VP97   | Q7vp97   | haemophilus  |
| 15 | 45   | 50.0 | 652  | 2 | Q6K967   | Q6k967   | oryza sativ  |
| 16 | 44   | 48.9 | 249  | 2 | Q9VSA8   | Q9vsa8   | drosophila   |
| 17 | 44   | 48.9 | 448  | 2 | Q9VM22   | Q9vm22   | drosophila   |
| 18 | 44   | 48.9 | 490  | 2 | Q8FE07   | Q8fe07   | escherichia  |
| 19 | 44   | 48.9 | 567  | 2 | Q6CVL7   | Q6cvl7   | kluveromyces |
| 20 | 44   | 48.9 | 595  | 2 | Q88AQ1   | Q88aq1   | pseudomonas  |
| 21 | 44   | 48.9 | 613  | 2 | Q6SHE5   | Q6she5   | uncultured   |
| 22 | 44   | 48.9 | 613  | 2 | AAR37676 | Aar37676 | unculture    |
| 23 | 44   | 48.9 | 754  | 2 | Q8T769   | Q8t769   | branchiosto  |
| 24 | 44   | 48.9 | 967  | 2 | Q8GZN4   | Q8gzn4   | lupinus alb  |
| 25 | 44   | 48.9 | 1609 | 2 | Q7XTW1   | Q7xtw1   | oryza sativ  |
| 26 | 43   | 47.8 | 161  | 2 | Q7X5A1   | Q7x5a1   | uncultured   |
| 27 | 43   | 47.8 | 201  | 1 | COX3_SYN | P50677   | synechococc  |
| 28 | 43   | 47.8 | 201  | 2 | Q8DHF0   | Q8dhf0   | synechococc  |
| 29 | 43   | 47.8 | 445  | 2 | O65259   | O65259   | arabidopsis  |
| 30 | 43   | 47.8 | 460  | 2 | Q8W495   | Q8w495   | arabidopsis  |
| 31 | 43   | 47.8 | 480  | 2 | Q740X0   | Q740x0   | mycobacteri  |
| 32 | 43   | 47.8 | 480  | 2 | AAS03539 | Aas03539 | mycobacte    |
| 33 | 43   | 47.8 | 904  | 2 | Q9NXZ1   | Q9nxz1   | homo sapien  |
| 34 | 43   | 47.8 | 1035 | 1 | DPOL_RH  | O71121   | rhesus cyto  |
| 35 | 43   | 47.8 | 1035 | 2 | AAP50613 | Aap50613 | rhesus cy    |
| 36 | 43   | 47.8 | 3076 | 2 | Q7PQY5   | Q7pqy5   | anopheles g  |
| 37 | 42.5 | 47.2 | 1285 | 2 | Q7XME3   | Q7xme3   | oryza sativ  |
| 38 | 42.5 | 47.2 | 1711 | 2 | Q7XS38   | Q7xs38   | oryza sativ  |
| 39 | 42   | 46.7 | 112  | 2 | Q8KKV4   | Q8kkv4   | rhizobium e  |
| 40 | 42   | 46.7 | 212  | 2 | Q7NK91   | Q7nk91   | gloeobacter  |
| 41 | 42   | 46.7 | 223  | 2 | Q9FKJ9   | Q9fkj9   | arabidopsis  |
| 42 | 42   | 46.7 | 271  | 2 | Q7UK83   | Q7uk83   | rhodospirill |
| 43 | 42   | 46.7 | 280  | 2 | Q7NIQ6   | Q7niq6   | gloeobacter  |
| 44 | 42   | 46.7 | 310  | 2 | O22768   | O22768   | arabidopsis  |
| 45 | 42   | 46.7 | 310  | 2 | Q8L9U3   | Q8l9u3   | arabidopsis  |

# ALIGNMENTS

## RESULT 1

Q7Q389

ID Q7Q389 PRELIMINARY; PRT; 341 AA.

AC Q7Q389;

DT 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26; Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE AgCP11001 (Fragment).

GN Name=agCG53822; ORFNames=ENSANGG00000007707;

OS Anopheles gambiae str. PEST.

OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;

OC Neoptera; Endopterygota; Diptera; Nematocera; Culicoidea; Anopheles.

OX NCBI\_TaxID=180454;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=PEST;

RA Anopheles Genome Sequencing Consortium;

RL Submitted (MAR-2002) to the EMBL/GenBank/DDBJ databases.

CC -!- CAUTION: The sequence shown here is derived from an

CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AAAB01008964; EAA12178.1; -.  
 DR InterPro; IPR006571; TLDc.  
 DR Pfam; PF07534; TLD; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 341 AA; 38864 MW; 8DD90A0CB268545E CRC64;

Query Match 56.7%; Score 51; DB 2; Length 341;  
 Best Local Similarity 75.0%; Pred. No. 7.2;  
 Matches 9; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 4 NLHQQTTPPDGFG 15  
 ||||| |:| |  
 Db 231 NLHQQTMPNGMG 242

# RESULT 2

## Q6MHZ1

ID Q6MHZ1 PRELIMINARY; PRT; 470 AA.  
 AC Q6MHZ1;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Rag3 (Two-component sensor histidine kinase) precursor.  
 GN Name=ragB; OrderedLocusNames=Bd3393;  
 OS Bdellovibrio bacteriovorus.  
 OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;  
 OC Bdellovibrionaceae; Bdellovibrio.  
 OX NCBI\_TaxID=959;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;  
 RX PubMed=14752164;  
 RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,  
 RA Keller H., Lambert C., Evans K.J., Goesmann A., Meyer F.,  
 RA Sockett R.E., Schuster S.C.;  
 RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus from a  
 RT genomic perspective.";  
 RL Science 303:689-692(2004).  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: Contains 1 histidine kinase domain.  
 DR EMBL; BX842655; CAE78191.1; -.  
 DR GO; GO:0016301; F:kinase activity; IEA.  
 DR InterPro; IPR003594; ATPbind\_ATPase.  
 DR InterPro; IPR004358; Bact\_sens\_pr\_C.  
 DR InterPro; IPR003660; HAMP.  
 DR InterPro; IPR005467; His\_kinase.  
 DR InterPro; IPR003661; His\_kinA\_N.  
 DR InterPro; IPR009082; His\_kin\_homodim.  
 DR Pfam; PF00672; HAMP; 1.  
 DR Pfam; PF02518; HATPase\_c; 1.  
 DR Pfam; PF00512; HisKA; 1.  
 DR PRINTS; PR00344; BCTRLSENSOR.  
 DR SMART; SM00304; HAMP; 1.  
 DR SMART; SM00387; HATPase\_c; 1.  
 DR SMART; SM00388; HisKA; 1.

DR PROSITE; PS50885; HAMP; 1.  
DR PROSITE; PS50109; HIS\_KIN; 1.  
KW Complete proteome; Kinase; Phosphorylation; Sensory transduction;  
KW Signal; Transferase; Transmembrane.  
FT SIGNAL 1 16 Potential.  
SQ SEQUENCE 470 AA; 53293 MW; D6B7DE9E6A41C68F CRC64;

Query Match 56.7%; Score 51; DB 2; Length 470;  
Best Local Similarity 69.2%; Pred. No. 10;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LLDNLHQQTTPPDG 13  
|||| |: ||| |  
Db 370 LLDNAHKYTPPGG 382

RESULT 3

CAE78191

ID CAE78191 PRELIMINARY; PRT; 470 AA.  
AC CAE78191;  
DT 02-MAR-2004 (TrEMBLrel. 27, Created)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last sequence update)  
DT 02-MAR-2004 (TrEMBLrel. 27, Last annotation update)  
DE RagB (Two-component sensor histidine kinase) precursor.  
GN RAGB OR BD3393.  
OS Bdellovibrio bacteriovorus.  
OC Bacteria; Proteobacteria; Deltaproteobacteria; Bdellovibrionales;  
OC Bdellovibrionaceae; Bdellovibrio.  
OX NCBI\_TaxID=959;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=HD100 / DSM 50701 / ATCC 15356 / NCIB 9529;  
RX PubMed=14752164;  
RA Rendulic S., Jagtap P., Rosinus A., Eppinger M., Baar C., Lanz C.,  
RA Keller H., Lambert C., Evans K.J., Goesmann A., Meyer F.,  
RA Sockett R.E., Schuster S.C.;  
RT "A predator unmasked: life cycle of Bdellovibrio bacteriovorus" from a  
RT genomic perspective.";  
RL Science 303:689-692(2004).  
DR EMBL; BX842655; CAE78191.1; -.  
KW Kinase; Signal.  
FT SIGNAL 1 16 Potential.  
SQ SEQUENCE 470 AA; 53293 MW; D6B7DE9E6A41C68F CRC64;

Query Match 56.7%; Score 51; DB 2; Length 470;  
Best Local Similarity 69.2%; Pred. No. 10;  
Matches 9; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LLDNLHQQTTPPDG 13  
|||| |: ||| |  
Db 370 LLDNAHKYTPPGG 382

RESULT 4

Q9UPJ7

ID Q9UPJ7 PRELIMINARY; PRT; 319 AA.  
AC Q9UPJ7;

DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE PkB-like (Fragment).  
 GN Name=PkB-like 2;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,  
 RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,  
 RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,  
 RA Ueng S., Tatum O., Campbell C., Fawcett J., Maltbie M., Misra M.,  
 RA Deaven L.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RA Ricke D.O.;  
 RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AC005591; AAC33798.1; -.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004672; F:protein kinase activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR InterPro; IPR011009; Kinase\_like.  
 DR InterPro; IPR011036; PH\_related.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 FT NON\_TER 1 1  
 SQ SEQUENCE 319 AA; 36620 MW; 01E18FFE1B5D4A53 CRC64;  
  
 Query Match 54.4%; Score 49; DB 2; Length 319;  
 Best Local Similarity 88.9%; Pred. No. 14;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
  
 Qy 3 DNLHQQTPP 11  
 :|||||||  
 Db 111 ENLHQQTPP 119

# RESULT 5

Q9UPJ8

ID Q9UPJ8 PRELIMINARY; PRT; 367 AA.  
 AC Q9UPJ8;  
 DT 01-MAY-2000 (TrEMBLrel. 13, Created)  
 DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE PkB-like (Fragment).  
 GN Name=PkB-like 1;  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]

RP SEQUENCE FROM N.A.  
RA Ricke D.O., Bruce D., Mundt M., Doggett N., Munk C., Saunders E.,  
RA Robinson D., Jones M., Buckingham J., Chasteen L., Thompson S.,  
RA Goodwin L., Bryant J., Tesmer J., Meincke L., Longmire J., White S.,  
RA Ueng S., Tatum O., Campbell C., Fawcett J., Maltbie M., Misra M.,  
RA Deaven L.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Ricke D.O.;  
RL Submitted (SEP-1998) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.  
DR EMBL; AC005591; AAC33797.1; -.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR InterPro; IPR011009; Kinase\_like.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; Pkinase; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
FT NON\_TER 1 1  
FT NON\_TER 367 367  
SQ SEQUENCE 367 AA; 41299 MW; 261CDF0075587493 CRC64;

Query Match 54.4%; Score 49; DB 2; Length 367;  
Best Local Similarity 88.9%; Pred. No. 17;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
:|||||||  
Db 340 ENLHQQTTP 348

# RESULT 6

## Q9BRD5

ID Q9BRD5 PRELIMINARY; PRT; 429 AA.  
AC Q9BRD5; Q8IV52;  
DT 01-JUN-2001 (TrEMBLrel. 17, Created)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)  
DT 01-OCT-2004 (TrEMBLrel. 28, Last annotation update)  
DE 3-phosphoinositide dependent protein kinase-1.  
GN Name=PDPK1;  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;

RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain, and Uterus;  
 RX MEDLINE=22388257; PubMed=12477932;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M., Butterfield Y.S.,  
 RA Krzywinski M.I., Skalska U., Smailus D.E., Schnerch A., Schein J.E.,  
 RA Jones S.J., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).

RN [2]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Uterus;  
 RA Strausberg R.;  
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.

RN [3]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RA Strausberg R.;  
 RL Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; BC006339; AAH06339.2; -.  
 DR EMBL; BC033494; AAH33494.1; -.  
 DR HSSP; O15530; 1H1W.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR Pfam; PF00069; Pkinase; 2.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 KW Kinase.  
 SQ SEQUENCE 429 AA; 48200 MW; 860C8A8C06161CE1 CRC64;

Query Match 54.4%; Score 49; DB 2; Length 429;  
 Best Local Similarity 88.9%; Pred. No. 20;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
 :|||||||  
 Db 221 ENLHQQTTP 229

RESULT 7

Q8K3L3

ID Q8K3L3 PRELIMINARY; PRT; 532 AA.  
 AC Q8K3L3;  
 DT 01-OCT-2002 (TrEMBLrel. 22, Created)  
 DT 01-OCT-2002 (TrEMBLrel. 22, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Phosphoinositide-dependent protein kinase-1 beta.  
 GN Name=Pdk1; Synonyms=Pdk1beta;  
 OS Mus musculus (Mouse).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
 OX NCBI\_TaxID=10090;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22050196; PubMed=12054753;  
 RA Dong L.Q., Ramos F.J., Wick M.J., Lim M.A., Guo Z., Strong R.,  
 RA Richardson A., Liu F.;  
 RT "Cloning and characterization of a testis and brain-specific isoform  
 RT of mouse 3'-phosphoinositide-dependent protein kinase-1, mPDK-1  
 RT beta.";  
 RL Biochem. Biophys. Res. Commun. 294:136-144(2002).  
 CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.  
 DR EMBL; AY062008; AAL47185.1; -.  
 DR HSSP; O15530; 1H1W.  
 DR MGD; MGI:1338068; Pdk1.  
 DR GO; GO:0016023; C:cytoplasmic vesicle; IDA.  
 DR GO; GO:0004676; F:3-phosphoinositide-dependent protein kinase. . .; IDA.  
 DR GO; GO:0006972; P:hyperosmotic response; IDA.  
 DR InterPro; IPR011009; Kinase\_like.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 532 AA; 60934 MW; F90731C7ECDEE589 CRC64;

Query Match 54.4%; Score 49; DB 2; Length 532;  
 Best Local Similarity 88.9%; Pred. No. 25;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTTP 11  
 :|||||||  
 Db 324 ENLHQQTTP 332

RESULT 8

Q810Z4



ID Q810Z4 PRELIMINARY; PRT; 551 AA.  
AC Q810Z4;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
DE PDK1 (Fragment).  
GN Name=Pdpk1;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=129/SvJ;  
RA Brathwaite M., Waeltz P., Schlessinger D., Nagaraja R.;  
RL Submitted (OCT-2002) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.  
DR EMBL; AY162410; AAO17164.1; -.  
DR GO; GO:0005524; F:ATP binding; IEA.  
DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
DR GO; GO:0004713; F:protein-tyrosine kinase activity; IEA.  
DR GO; GO:0016740; F:transferase activity; IEA.  
DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
DR InterPro; IPR011009; Kinase\_like.  
DR InterPro; IPR000719; Prot\_kinase.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
DR InterPro; IPR001245; Tyr\_pkinase.  
DR Pfam; PF00069; Pkinase; 1.  
DR PRINTS; PR00109; TYRKINASE.  
DR ProDom; PD000001; Prot\_kinase; 1.  
DR SMART; SM00220; S\_TKc; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
FT NON\_TER 1 1  
SQ SEQUENCE 551 AA; 62869 MW; ACC31D51439282F4 CRC64;

Query Match 54.4%; Score 49; DB 2; Length 551;  
Best Local Similarity 88.9%; Pred. No. 26;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTPP 11  
:|||||||  
Db 343 ENLHQQTPP 351

#### RESULT 9

##### PDPK\_HUMAN

ID PDPK\_HUMAN STANDARD; PRT; 556 AA.  
AC C15530;  
DT 16-OCT-2001 (Rel. 40, Created).  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 01-OCT-2004 (Rel. 45, Last annotation update)  
DE 3-phosphoinositide dependent protein kinase-1 (EC 2.7.1.37) (hPDK1).  
GN Name=PDPK1; Synonyms=PDK1;

OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=97250749; PubMed=9094314;  
 RA Alessi D.R., James S.R., Downes C.P., Holmes A.B., Gaffney P.R.J.,  
 RA Reese C.B., Cohen P.;  
 RT "Characterization of a 3-phosphoinositide-dependent protein kinase  
 RT which phosphorylates and activates protein kinase B alpha.";  
 RL Curr. Biol. 7:261-269(1997).  
 RN [2]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RX MEDLINE=98035195; PubMed=9368760;  
 RA Alessi D.R., Deak M., Casamayor A., Caudwell F.B., Morrice N.A.,  
 RA Norman D.G., Gaffney P.R.J., Reese C.B., MacDougall C.N., Harbison D.,  
 RA Ashworth A., Bownes M.;  
 RT "3-phosphoinositide-dependent protein kinase-1 (PDK1): structural and  
 RT functional homology with the Drosophila DSTPK61 kinase.";  
 RL Curr. Biol. 7:776-789(1997).  
 RN [3]  
 RP SEQUENCE FROM N.A. (ISOFORMS 1 AND 3).  
 RC TISSUE=Myeloid;  
 RX MEDLINE=98111410; PubMed=9445477;  
 RA Stephens L.R., Anderson K.E., Stokoe D., Erdjument-Bromage H.,  
 RA Painter G.F., Holmes A.B., Gaffney P.R.J., Reese C.B., McCormick F.,  
 RA Tempst P., Coadwell W.J., Hawkins P.T.;  
 RT "Protein kinase B kinases that mediate phosphatidylinositol 3,4,5-  
 RT trisphosphate-dependent activation of protein kinase B.";  
 RL Science 279:710-714(1998).  
 RN [4]  
 RP SEQUENCE FROM N.A. (ISOFORM 1).  
 RC TISSUE=Kidney;  
 RX MEDLINE=22388257; PubMed=12477932; DOI=10.1073/pnas.242603899;  
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,  
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,  
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,  
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,  
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,  
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,  
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,  
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,  
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,  
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,  
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,  
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,  
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,  
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,  
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,  
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,  
 RA Schnierch A., Schein J.E., Jones S.J.M., Marra M.A.;  
 RT "Generation and initial analysis of more than 15,000 full-length human  
 RT and mouse cDNA sequences.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).  
 RN [5]  
 RP MUTAGENESIS OF ARG-474, AND ALTERNATIVE SPLICING.

RX MEDLINE=98301766; PubMed=9637919;  
 RA Anderson K.E., Coadwell W.J., Stephens L.R., Hawkins P.T.;  
 RT "Translocation of PDK-1 to the plasma membrane is important in  
 RT allowing PDK-1 to activate protein kinase B.";  
 RL Curr. Biol. 8:684-691(1998).  
 RN [6]  
 RP PHOSPHORYLATION SITES SER-25; SER-241; SER-393; SER-396 AND SER-410,  
 RP AND MUTAGENESIS OF SER-25; SER-241; SER-393; SER-396 AND SER-410.  
 RX MEDLINE=99386657; PubMed=10455013;  
 RA Casamayor A., Morrice N.A., Alessi D.R.;  
 RT "Phosphorylation of Ser-241 is essential for the activity of 3-  
 RT phosphoinositide-dependent protein kinase-1: identification of five  
 RT sites of phosphorylation in vivo.";  
 RL Biochem. J. 342:287-292(1999).  
 RN [7]  
 RP PHOSPHORYLATION SITES TYR-9; SER-241; TYR-373 AND TYR-376, AND  
 RP MUTAGENESIS OF TYR-9; TYR-373 AND TYR-376.  
 RX MEDLINE=21463095; PubMed=11481331; DOI=10.1074/jbc.M105916200;  
 RA Park J., Hill M.M., Hess D., Brazil D.P., Hofsteenge J.,  
 RA Hemmings B.A.;  
 RT "Identification of tyrosine phosphorylation sites on 3-  
 RT phosphoinositide-dependent protein kinase-1 (PDK1) and their role in  
 RT regulating kinase activity.";  
 RL J. Biol. Chem. 276:37459-37471(2001).  
 CC -!- FUNCTION: Phosphorylates and activates not only PKB/AKT, but also  
 CC PKA, PKC-zeta, p70S6K and p90S6K/RSK. May play a general role in  
 CC signaling processes and in development (By similarity). Isoform 3  
 CC is catalytically inactive.  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated after  
 CC cell stimulation leading to its translocation. Tyrosine  
 CC phosphorylation seems to occur only at the plasma membrane.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=3;  
 CC Name=1;  
 CC IsoId=O15530-1; Sequence=Displayed;  
 CC Name=2;  
 CC IsoId=O15530-2; Sequence=VSP\_004894;  
 CC Name=3;  
 CC IsoId=O15530-3; Sequence=VSP\_004895;  
 CC -!- TISSUE SPECIFICITY: Appears to be expressed ubiquitously.  
 CC -!- PTM: Phosphorylated on tyrosine and serine/threonine.  
 CC Phosphorylation on Ser-241 in the activation loop is required for  
 CC full activity. PDK1 itself can autophosphorylate Ser-241, leading  
 CC to its own activation.  
 CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family. PDK1  
 CC subfamily.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----

DR EMBL; AF017995; AAC51825.1; -.  
 DR EMBL; Y15056; CAA75341.1; -.  
 DR EMBL; BC012103; AAH12103.1; -.  
 DR PDB; 1H1W; X-ray; A=71-359.  
 DR Genew; HGNC:8816; PDPK1.  
 DR MIM; 605213; -.  
 DR GO; GO:0005737; C:cytoplasm; IEP.  
 DR GO; GO:0005886; C:plasma membrane; IEP.  
 DR GO; GO:0004676; F:3-phosphoinositide-dependent protein kinase. . .; TAS.  
 DR GO; GO:0030036; P:actin cytoskeleton organization and biogenesis; TAS.  
 DR GO; GO:0008286; P:insulin receptor signaling pathway; TAS.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; TAS.  
 DR InterPro; IPR011009; Kinase\_like.  
 DR InterPro; IPR001849; PH.  
 DR InterPro; IPR011036; PH\_related.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR002290; Ser\_thr\_pkinase.  
 DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
 DR InterPro; IPR001245; Tyr\_pkinase.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR PRINTS; PR00109; TYRKINASE.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR SMART; SM00220; S\_TKc; 1.  
 DR PROSITE; PS50003; PH\_DOMAIN; FALSE\_NEG.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW 3D-structure; Alternative splicing; ATP-binding; Membrane;  
 KW Phosphorylation; Serine/threonine-protein kinase; Transferase.  
 FT DOMAIN 82 342 Protein kinase.  
 FT DOMAIN 459 550 PH.  
 FT NP\_BIND 88 96 ATP (By similarity).  
 FT BINDING 111 111 ATP (By similarity).  
 FT ACT\_SITE 205 205 Proton acceptor (By similarity).  
 FT DOMAIN 389 398 Poly-Ser.  
 FT MOD\_RES 9 9 Phosphotyrosine.  
 FT MOD\_RES 25 25 Phosphoserine.  
 FT MOD\_RES 241 241 Phosphoserine (by autocatalysis).  
 FT MOD\_RES 373 373 Phosphotyrosine.  
 FT MOD\_RES 376 376 Phosphotyrosine.  
 FT MOD\_RES 393 393 Phosphoserine.  
 FT MOD\_RES 396 396 Phosphoserine.  
 FT MOD\_RES 410 410 Phosphoserine.  
 FT VARSPLIC 1 50 Missing (in isoform 2).  
 FT VARSPLIC 238 263 /FTId=VSP\_004894.  
 FT VARSPLIC 238 263 Missing (in isoform 3).  
 FT VARSPLIC 238 263 /FTId=VSP\_004895.  
 FT MUTAGEN 9 9 Y->F: Slight reduction in pervanadate-stimulated tyrosine phosphorylation.  
 FT MUTAGEN 25 25 S->A: No effect.  
 FT MUTAGEN 241 241 S->A: No activation.  
 FT MUTAGEN 393 393 S->A: No effect.  
 FT MUTAGEN 396 396 S->A: No effect.  
 FT MUTAGEN 410 410 S->A: No effect.  
 FT MUTAGEN 474 474 R->A: No PDGF-dependent translocation to the membrane.  
 FT MUTAGEN 373 373 Y->F: Reduction in basal activity.

FT MUTAGEN 376 376 Y->F: Reduction in basal activity.  
SQ SEQUENCE 556 AA; 63151 MW; ED8C0306DC4D0653 CRC64;

Query Match 54.4%; Score 49; DB 1; Length 556;  
Best Local Similarity 88.9%; Pred. No. 26;  
Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 3 DNLHQQTPP 11  
:|||||||  
Db 348 ENLHQQTPP 356

#### RESULT 10

##### PDPK\_MOUSE

ID PDPK\_MOUSE STANDARD; PRT; 559 AA.  
AC Q9Z2A0; Q9R1D8; Q9R215;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE 3-phosphoinositide dependent protein kinase-1 (EC 2.7.1.37) (mPDK1).  
GN Name=Pdk1; Synonyms=Pdk1;  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Liver;  
RX MEDLINE=99175193; PubMed=10075713;  
RA Dong L.Q., Zhang R.-B., Langlais P., He H., Clark M., Zhu L., Liu F.;  
RT "Primary structure, tissue distribution, and expression of mouse  
RT phosphoinositide-dependent protein kinase-1, a protein kinase that  
RT phosphorylates and activates protein kinase C zeta."  
RL J. Biol. Chem. 274:8117-8122 (1999).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC TISSUE=Brain;  
RA Park J., Hemmings B.A.;  
RT "Mouse phosphoinositide-dependent protein kinase 1 (mPDK1).";  
RL Submitted (FEB-1999) to the EMBL/GenBank/DDBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN=C57BL/6;  
RA Xu P., Taylor S.;  
RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.  
CC -!- FUNCTION: Phosphorylates and activates not only PKB/AKT, but also  
CC PKA, PKC-zeta, p70S6K and p90S6K/RSK. May play a general role in  
CC signaling processes and in development. Could also play a role in  
CC sex differentiation processes.  
CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
CC -!- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated after  
CC cell stimulation leading to its translocation. Tyrosine  
CC phosphorylation seems to occur only at the plasma membrane.  
CC -!- TISSUE SPECIFICITY: Highly expressed in heart, brain, liver and  
CC testis, also expressed in embryonic cells.  
CC -!- PTM: Phosphorylated on tyrosine and serine/threonine.  
CC Phosphorylation on Ser-244 in the activation loop is required for

```

CC      full activity. PDK1 itself can autophosphorylate Ser-244, leading
CC      to its own activation (By similarity).
CC      -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family. PDK1
CC      subfamily.
CC      -!- SIMILARITY: Contains 1 PH domain.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AF086625; AAC67544.1; -.
DR      EMBL; AF126294; AAD38505.1; -.
DR      EMBL; AF079535; AAC96115.1; -.
DR      HSSP; O15530; 1H1W.
DR      MGD; MGI:1338068; Pdkp1.
DR      GO; GO:0004676; F:3-phosphoinositide-dependent protein kinase. . .; IDA.
DR      InterPro; IPR011009; Kinase_like.
DR      InterPro; IPR001849; PH.
DR      InterPro; IPR000719; Prot_kinase.
DR      InterPro; IPR002290; Ser_thr_pkinase.
DR      InterPro; IPR008271; Ser_thr_pkin_AS.
DR      InterPro; IPR001245; Tyr_pkinase.
DR      Pfam; PF00069; Pkinase; 1.
DR      PRINTS; PR00109; TYRKINASE.
DR      ProDom; PD000001; Prot_kinase; 1.
DR      SMART; SM00220; S_TKc; 1.
DR      PROSITE; PS50003; PH_DOMAIN; FALSE_NEG.
DR      PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR      PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR      PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW      ATP-binding; Membrane; Phosphorylation;
KW      Serine/threonine-protein kinase; Transferase.
FT      DOMAIN      85      345      Protein kinase.
FT      DOMAIN      462      553      PH.
FT      NP_BIND      91      99      ATP (By similarity).
FT      BINDING      114      114      ATP (By similarity).
FT      ACT_SITE      208      208      Proton acceptor (By similarity).
FT      DOMAIN      392      401      Poly-Ser.
FT      MOD_RES      9      9      Phosphotyrosine (By similarity).
FT      MOD_RES      25      25      Phosphoserine (By similarity).
FT      MOD_RES      244      244      Phosphoserine (by autocatalysis) (By
FT                                     similarity).
FT      MOD_RES      376      376      Phosphotyrosine (By similarity).
FT      MOD_RES      379      379      Phosphotyrosine (By similarity).
FT      MOD_RES      396      396      Phosphoserine (By similarity).
FT      MOD_RES      399      399      Phosphoserine (By similarity).
FT      MOD_RES      406      406      Phosphoserine (By similarity).
FT      MOD_RES      413      413      Phosphoserine (By similarity).
FT      CONFLICT      84      84      D -> N (in Ref. 1).
FT      CONFLICT      248      248      T -> P (in Ref. 3).
FT      CONFLICT      285      285      F -> S (in Ref. 3).
FT      CONFLICT      546      546      W -> R (in Ref. 3).
SQ      SEQUENCE      559 AA; 63758 MW; F2A617A27460FAC9 CRC64;

```

Query Match 54.4%; Score 49; DB 1; Length 559;  
 Best Local Similarity 88.9%; Pred. No. 27;  
 Matches 8; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 3 DNLHQQTPP 11  
 :|||||||  
 Db 351 ENLHQQTPP 359

RESULT 11

PDPK\_RAT

ID PDPK\_RAT STANDARD; PRT; 559 AA.  
 AC O55173;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE 3-phosphoinositide dependent protein kinase-1 (EC 2.7.1.37) (Protein  
 DE kinase B kinase) (Pkb kinase).  
 GN Name=Pdk1; Synonyms=Pdk1;  
 OS Rattus norvegicus (Rat).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
 OX NCBI\_TaxID=10116;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Brain;  
 RX MEDLINE=98111410; PubMed=9445477;  
 RA Stephens L., Anderson K.E., Stokoe D., Erdjument-Bromage H.,  
 RA Painter G.F., Holmes A.B., Gaffney P.R.J., Reese C.B., McCormick F.,  
 RA Tempst P., Coadwell W.J., Hawkins P.T.;  
 RT "Protein kinase B kinases that mediate phosphatidylinositol 3,4,5-  
 RT trisphosphate-dependent activation of protein kinase B."  
 RL Science 279:710-714 (1998).  
 CC -!- FUNCTION: Phosphorylates and activates not only PKB/AKT, but also  
 CC PKA, PKC-zeta, p70S6K and p90S6K/RSK. May play a general role in  
 CC signaling processes and in development (By similarity).  
 CC -!- CATALYTIC ACTIVITY: ATP + a protein = ADP + a phosphoprotein.  
 CC -!- SUBCELLULAR LOCATION: Cytoplasmic and membrane-associated after  
 CC cell stimulation leading to its translocation. Tyrosine  
 CC phosphorylation seems to occur only at the plasma membrane (By  
 CC similarity).  
 CC -!- PTM: Phosphorylated on tyrosine and serine/threonine.  
 CC Phosphorylation on Ser-244 in the activation loop is required for  
 CC full activity. PDK1 itself can autophosphorylate Ser-244, leading  
 CC to its own activation (By similarity).  
 CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family. PDK1  
 CC subfamily.  
 CC -!- SIMILARITY: Contains 1 PH domain.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

```

CC -----
DR EMBL; Y15748; CAA75758.1; -.
DR HSSP; O15530; 1H1W.
DR RGD; 620307; Pdpk1.
DR InterPro; IPR011009; Kinase_like.
DR InterPro; IPR001849; PH.
DR InterPro; IPR000719; Prot_kinase.
DR InterPro; IPR002290; Ser_thr_pkinase.
DR InterPro; IPR008271; Ser_thr_pkin_AS.
DR InterPro; IPR001245; Tyr_pkinase.
DR Pfam; PF00069; Pkinase; 1.
DR PRINTS; PR00109; TYRKINASE.
DR ProDom; PD000001; Prot_kinase; 1.
DR SMART; SM00220; S_TKc; 1.
DR PROSITE; PS50003; PH_DOMAIN; FALSE_NEG.
DR PROSITE; PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE; PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE; PS00108; PROTEIN_KINASE_ST; 1.
KW ATP-binding; Membrane; Phosphorylation;
KW Serine/threonine-protein kinase; Transferase.
FT DOMAIN      85      345      Protein kinase.
FT DOMAIN      462     553      PH.
FT NP_BIND      91      99      ATP (By similarity).
FT BINDING     114     114      ATP (By similarity).
FT ACT_SITE    208     208      Proton acceptor (By similarity).
FT DOMAIN      392     399      Poly-Ser.
FT MOD_RES      9       9      Phosphotyrosine (By similarity).
FT MOD_RES     25      25      Phosphoserine (By similarity).
FT MOD_RES    244     244      Phosphoserine (by autocatalysis) (By
FT                                     similarity).
FT MOD_RES    376     376      Phosphotyrosine (By similarity).
FT MOD_RES    379     379      Phosphotyrosine (By similarity).
FT MOD_RES    396     396      Phosphoserine (By similarity).
FT MOD_RES    399     399      Phosphoserine (By similarity).
FT MOD_RES    406     406      Phosphoserine (By similarity).
FT MOD_RES    413     413      Phosphoserine (By similarity).
SQ SEQUENCE    559 AA;  63609 MW;  ADE70A7F6C2A20BF CRC64;

```

```

Query Match          54.4%;  Score 49;  DB 1;  Length 559;
Best Local Similarity 88.9%;  Pred. No. 27;
Matches      8;  Conservative 1;  Mismatches 0;  Indels 0;  Gaps 0;

```

```

QY      3 DNLHQQTPP 11
      :|||
Db      351 ENLHQQTPP 359

```

```

RESULT 12
CBP1_CAEEL
ID CBP1_CAEEL STANDARD; PRT; 2056 AA.
AC P34545;
DT 01-FEB-1994 (Rel. 28, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 05-JUL-2004 (Rel. 44, Last annotation update)
DE Protein cbp-1.
GN Name=cbp-1; ORFNames=R10E11.1;
OS Caenorhabditis elegans.

```



OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
 OC Rhabditidae; Peloderinae; Caenorhabditis.  
 OX NCBI\_TaxID=6239;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RX MEDLINE=94150718; PubMed=7906398;  
 RA Wilson R., Ainscough R., Anderson K., Baynes C., Berks M.,  
 RA Bonfield J., Burton J., Connell M., Copsey T., Cooper J., Coulson A.,  
 RA Craxton M., Dear S., Du Z., Durbin R., Favello A., Fraser A.,  
 RA Fulton L., Gardner A., Green P., Hawkins T., Hillier L., Jier M.,  
 RA Johnston L., Jones M., Kershaw J., Kirsten J., Laissner N.,  
 RA Latreille P., Lightning J., Lloyd C., Mortimore B., O'Callaghan M.,  
 RA Parsons J., Percy C., Rifken L., Roopra A., Saunders D., Shownkeen R.,  
 RA Sims M., Smaldon N., Smith A., Smith M., Sonnhammer E., Staden R.,  
 RA Sulston J., Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K.,  
 RA Waterston R., Watson A., Weinstock L., Wilkinson-Sproat J.,  
 RA Wohldman P.;  
 RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of *C.*  
 RT *elegans*.";   
 RL Nature 368:32-38 (1994).  
 RN [2]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RA Durbin R.;  
 RL Submitted (DEC-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=b;  
 CC IsoId=P34545-1; Sequence=Displayed;  
 CC Name=a;  
 CC IsoId=P34545-2; Sequence=VSP\_000557;  
 CC Note=No experimental confirmation available;  
 CC -!- SIMILARITY: Contains 1 bromodomain.  
 CC -!- SIMILARITY: Contains 1 KIX domain.  
 CC -!- SIMILARITY: Contains 2 TAZ-type zinc fingers.  
 CC -!- SIMILARITY: Contains 1 ZZ-type zinc finger.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; Z29095; CAA82353.2; -.  
 DR EMBL; Z29095; CAD18875.1; -.  
 DR PIR; G88564; G88564.  
 DR HSSP; P45481; 1L8C.  
 DR WormPep; R10E11.1a; CE28069.  
 DR WormPep; R10E11.1b; CE21117.  
 DR InterPro; IPR001487; Bromodomain.  
 DR InterPro; IPR010303; DUF902.  
 DR InterPro; IPR009255; DUF906.  
 DR InterPro; IPR003101; KIX.  
 DR InterPro; IPR000197; TAZ\_finger.  
 DR InterPro; IPR001965; Znf\_PHD.

DR InterPro; IPR000433; Znf\_ZZ.  
 DR Pfam; PF00439; Bromodomain; 1.  
 DR Pfam; PF06001; DUF902; 1.  
 DR Pfam; PF06010; DUF906; 1.  
 DR Pfam; PF02172; KIX; 1.  
 DR Pfam; PF02135; zf-TAZ; 2.  
 DR Pfam; PF00569; ZZ; 1.  
 DR PRINTS; PR00503; BROMODOMAIN.  
 DR SMART; SM00297; BROMO; 1.  
 DR SMART; SM00551; Znf\_TAZ; 2.  
 DR SMART; SM00291; Znf\_ZZ; 1.  
 DR PROSITE; PS00633; BROMODOMAIN\_1; 1.  
 DR PROSITE; PS50014; BROMODOMAIN\_2; 1.  
 DR PROSITE; PS50952; KIX; 1.  
 DR PROSITE; PS01359; ZF\_PHD\_1; 1.  
 DR PROSITE; PS50134; ZF\_TAZ; 2.  
 DR PROSITE; PS01357; ZF\_ZZ\_1; 1.  
 DR PROSITE; PS50135; ZF\_ZZ\_2; 1.  
 KW Alternative splicing; Bromodomain; Metal-binding; Repeat; Zinc;  
 KW Zinc-finger.  
 FT ZN\_FING 399 505 TAZ-type 1.  
 FT DOMAIN 593 672 KIX.  
 FT DOMAIN 881 953 Bromodomain.  
 FT ZN\_FING 1493 1534 ZZ-type.  
 FT ZN\_FING 1550 1631 TAZ-type 2.  
 FT DOMAIN 1687 2008 GLY/GLN-RICH.  
 FT VARSPLIC 467 478 SDTTQTTKKCSV -> F (in isoform a).  
 FT /FTid=VSP\_000557.  
 SQ SEQUENCE 2056 AA; 227179 MW; 949FF4608C634F01 CRC64;

Query Match 51.1%; Score 46; DB 1; Length 2056;  
 Best Local Similarity 70.0%; Pred. No. 3.5e+02;  
 Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

QY 4 NLHQQTTPPDG 13  
 |:||| ||:|  
 Db 735 NMHQQIPPNG 744

# RESULT 13

Q7U666

ID Q7U666 PRELIMINARY; PRT; 183 AA.  
 AC Q7U666;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Hypothetical.  
 GN OrderedLocusNames=SYNW1473;  
 OS Synechococcus sp. (strain WH8102).  
 OC Bacteria; Cyanobacteria; Chroococcales; Synechococcus.  
 OX NCBI\_TaxID=84588;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22825697; PubMed=12917641; DOI=10.1038/nature01943;  
 RA Palenik B., Brahamsha B., Larimer F.W., Land M.L., Hauser L.,  
 RA Chain P., Lamerdin J.E., Regala W., Allen E.E., McCarren J.,  
 RA Paulsen I.T., Dufresne A., Partensky F., Webb E.A., Waterbury J.;

RT "The genome of a motile marine Synechococcus.";  
RL Nature 424:1037-1042(2003).  
DR EMBL; BX569693; CAE07988.1; -.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 183 AA; 21125 MW; 3F1532EF5C1E2FF4 CRC64;

Query Match 50.0%; Score 45; DB 2; Length 183;  
Best Local Similarity 66.7%; Pred. No. 36;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 2 LDNLHQQTTPPDG 13  
| | | | : | | |  
Db 64 LGNLHRWLPPDG 75

#### RESULT 14

Q7VP97

ID Q7VP97 PRELIMINARY; PRT; 332 AA.  
AC Q7VP97;  
DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocusNames=HD0195;  
OS Haemophilus ducreyi.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pasteurellales;  
OC Pasteurellaceae; Haemophilus.  
OX NCBI\_TaxID=730;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=35000HP / ATCC 700724;  
RA Munson R.S. Jr., Ray W.C., Mahairas G., Sabo P., Mungur R.,  
RA Johnson L., Nguyen D., Wang J., Forst C., Hood L.;  
RT "The complete genome sequence of Haemophilus ducreyi.";  
RL Submitted (JUN-2003) to the EMBL/GenBank/DDBJ databases.  
DR EMBL; AE017151; AAP95188.1; -.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 332 AA; 37655 MW; B693E0C0673577E8 CRC64;

Query Match 50.0%; Score 45; DB 2; Length 332;  
Best Local Similarity 53.3%; Pred. No. 69;  
Matches 8; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 LLDNLHQQTTPPDGFG 15  
| | : : | : | | : |  
Db 309 LLNLAYQRTPKDGYG 323

#### RESULT 15

Q6K967

ID Q6K967 PRELIMINARY; PRT; 652 AA.  
AC Q6K967;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Putative hexose transporter.  
GN Name=OJ1149\_C12.19;

OS Oryza sativa (japonica cultivar-group).  
 OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
 OC Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
 OC Ehrhartoideae; Oryzeae; Oryza.  
 OX NCBI\_TaxID=39947;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Sasaki T., Matsumoto T., Yamamoto K.;  
 RL Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.  
 CC -!- SUBCELLULAR LOCATION: Integral membrane protein (By similarity).  
 CC -!- SIMILARITY: Belongs to the sugar transporter family.  
 DR EMBL; AP004082; BAD23011.1; -.  
 DR InterPro; IPR000566; Lipocln\_cytFABP.  
 DR InterPro; IPR007114; MFS.  
 DR InterPro; IPR005828; Sub\_transporter.  
 DR InterPro; IPR003663; Sugar\_transpt.  
 DR InterPro; IPR005829; Sug\_transporter.  
 DR Pfam; PF00083; Sugar\_tr; 1.  
 DR PRINTS; PR00171; SUGRTRNSPORT.  
 DR PROSITE; PS00213; LIPOCALIN; UNKNOWN\_1.  
 DR PROSITE; PS50850; MFS; 1.  
 DR PROSITE; PS00217; SUGAR\_TRANSPORT\_2; 1.  
 KW Sugar transport; Transmembrane; Transport.  
 SQ SEQUENCE 652 AA; 68827 MW; EEE20446D2F9B1F6 CRC64;

Query Match 50.0%; Score 45; DB 2; Length 652;  
 Best Local Similarity 61.5%; Pred. No. 1.4e+02;  
 Matches 8; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 LLDNLHQQTTPPDG 13  
 |||:|| |||  
 Db 287 LLDSLHDMNPPAG 299

Search completed: January 31, 2005, 13:22:43  
 Job time : 124.818 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 12:55:49 ; Search time 91.9545 Seconds  
(without alignments)  
54.616 Million cell updates/sec

Title: US-10-067-620-8  
Perfect score: 80  
Sequence: 1 YSDGNFFGAGLDHQ 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2002273 seqs, 358729299 residues

Total number of hits satisfying chosen parameters: 2002273

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_23Sep04:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query |        | DB | ID       | Description        |
|------------|-------|---------|--------|----|----------|--------------------|
|            |       | Match   | Length |    |          |                    |
| 1          | 80    | 100.0   | 14     | 5  | ABB81975 | Abb81975 30 kDa ra |
| 2          | 43    | 53.8    | 851    | 7  | ABO66766 | Abo66766 Klebsiell |
| 3          | 42    | 52.5    | 513    | 6  | ABU18793 | Abu18793 Protein e |
| 4          | 41    | 51.2    | 171    | 2  | AAW23689 | Aaw23689 Potato po |
| 5          | 41    | 51.2    | 215    | 3  | AA70004  | Aay70004 Casein ki |
| 6          | 41    | 51.2    | 215    | 4  | ABB66949 | Abb66949 Drosophil |
| 7          | 41    | 51.2    | 215    | 4  | ABB66950 | Abb66950 Drosophil |
| 8          | 41    | 51.2    | 215    | 4  | ABB70309 | Abb70309 Drosophil |
| 9          | 41    | 51.2    | 215    | 5  | ABG79668 | Abg79668 Human cas |

|    |    |      |     |   |          |          |           |
|----|----|------|-----|---|----------|----------|-----------|
| 10 | 41 | 51.2 | 215 | 5 | ABG79669 | Abg79669 | Mouse cas |
| 11 | 41 | 51.2 | 215 | 7 | ADB79767 | Adb79767 | Rat Casei |
| 12 | 41 | 51.2 | 215 | 7 | ADC53769 | Adc53769 | Casein ki |
| 13 | 41 | 51.2 | 215 | 7 | ADE57086 | Ade57086 | Human Pro |
| 14 | 41 | 51.2 | 215 | 7 | ADE57088 | Ade57088 | Rat Prote |
| 15 | 41 | 51.2 | 215 | 7 | ADD47777 | Add47777 | Human Pro |
| 16 | 41 | 51.2 | 215 | 7 | ADD47775 | Add47775 | Rat Prote |
| 17 | 41 | 51.2 | 215 | 7 | ADE57084 | Ade57084 | Rat Prote |
| 18 | 41 | 51.2 | 215 | 7 | ADE57090 | Ade57090 | Human Pro |
| 19 | 41 | 51.2 | 215 | 8 | ADK60204 | Adk60204 | Angiogene |
| 20 | 41 | 51.2 | 215 | 8 | ADK60420 | Adk60420 | Angiogene |
| 21 | 41 | 51.2 | 215 | 8 | ADJ57058 | Adj57058 | Human cas |
| 22 | 41 | 51.2 | 215 | 8 | ADK60505 | Adk60505 | Angiogene |
| 23 | 41 | 51.2 | 215 | 8 | ADK60721 | Adk60721 | Angiogene |
| 24 | 41 | 51.2 | 215 | 8 | ADL72083 | Adl72083 | Human cas |
| 25 | 41 | 51.2 | 215 | 8 | ADL72085 | Adl72085 | Rat casei |
| 26 | 41 | 51.2 | 215 | 8 | ADO57527 | Ado57527 | Human CKI |
| 27 | 41 | 51.2 | 215 | 8 | ADP73128 | Adp73128 | Angiogene |
| 28 | 41 | 51.2 | 215 | 8 | ADP73344 | Adp73344 | Human cas |
| 29 | 41 | 51.2 | 215 | 8 | ADP74599 | Adp74599 | Amino aci |
| 30 | 41 | 51.2 | 219 | 4 | ABB64071 | Abb64071 | Drosophil |
| 31 | 41 | 51.2 | 223 | 2 | AAW97991 | Aaw97991 | Tobacco p |
| 32 | 41 | 51.2 | 223 | 3 | AAB13227 | Aab13227 | Ascoris s |
| 33 | 41 | 51.2 | 227 | 2 | AAW97992 | Aaw97992 | Tobacco p |
| 34 | 41 | 51.2 | 269 | 5 | ABP41834 | Abp41834 | Human ova |
| 35 | 41 | 51.2 | 379 | 3 | AAB18570 | Aab18570 | Amino aci |
| 36 | 41 | 51.2 | 438 | 2 | AAR33772 | Aar33772 | Potato tu |
| 37 | 41 | 51.2 | 456 | 4 | ABB52485 | Abb52485 | Escherich |
| 38 | 41 | 51.2 | 464 | 7 | ADC01568 | Adc01568 | Enterohae |
| 39 | 41 | 51.2 | 596 | 2 | AAR39554 | Aar39554 | Deduced a |
| 40 | 41 | 51.2 | 596 | 4 | AAB30862 | Aab30862 | Amino aci |
| 41 | 40 | 50.0 | 103 | 3 | AAG54760 | Aag54760 | Arabidops |
| 42 | 40 | 50.0 | 131 | 2 | AAY34149 | Aay34149 | Human tru |
| 43 | 40 | 50.0 | 133 | 3 | AAG61694 | Aag61694 | Arabidops |
| 44 | 40 | 50.0 | 216 | 3 | AAG05723 | Aag05723 | Arabidops |
| 45 | 40 | 50.0 | 216 | 8 | ADN74423 | Adn74423 | Thale cre |

# ALIGNMENTS

## RESULT 1

ABB81975

ID ABB81975 standard; peptide; 14 AA.

XX

AC ABB81975;

XX

DT 25-NOV-2002 (first entry)

XX

DE 30 kDa ragweed pollen allergen tryptic peptide 8.

XX

KW Ragweed; pollen; allergen; Ambt 7; glycoprotein; antiallergic;  
KW immunotherapy; disulphide protein.

XX

OS Ambrosia elatior.

XX

PN WO200263012-A2.

XX  
PD 15-AUG-2002.  
XX  
PF 04-FEB-2002; 2002WO-US003346.  
XX  
PR 05-FEB-2001; 2001US-0266686P.  
XX  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Buchanan BB, Del Val G, Frick OL;  
XX  
DR WPI; 2002-657539/70.  
XX  
PT New ragweed pollen allergens, useful in allergy testing and immunotherapy  
PT regimens, particularly for treating sensitivity to pollen or pollen  
PT allergy (e.g. anaphylaxis, or symptoms of hives or asthma) in a mammal,  
PT especially a human.  
XX  
PS Claim 1; Page 53; 70pp; English.  
XX  
CC The invention relates to an isolated pollen allergen purified from  
CC ragweed pollen, substantially free of any other pollen proteins, or a  
CC protein that is an antigenic fragment of a pollen allergen Ambt 7. The  
CC allergen is characterized by the following physiochemical and biological  
CC properties: (a) being contained in pollen extracts; (b) a glycoprotein;  
CC (c) a sulphhydryl group containing protein; (d) a molecular weight of  
CC about 30 kDa as determined by SDS-polyacrylamide gel electrophoresis; and  
CC (e) possessing allergen activity. The pollen allergen, or antigenic  
CC protein fragment of the pollen allergen Ambt 7, or composition is useful  
CC for treating sensitivity to pollen or pollen allergy in a mammal. This  
CC allergy includes anaphylaxis or atopy, which includes the symptoms of hay  
CC fever, asthma or hives. The allergen is also useful in allergy testing  
CC and immunotherapy regimens. Sequences ABB81968-978 represent tryptic  
CC peptide fragments of the 30 kDa ragweed complete pollen extract  
CC disulphide protein allergen  
XX  
SQ Sequence 14 AA;

Query Match 100.0%; Score 80; DB 5; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.8e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDHQ 14  
|||||||  
Db 1 YSDGNFFGAGLDHQ 14

RESULT 2  
ABO66766  
ID ABO66766 standard; protein; 851 AA.  
XX  
AC ABO66766;  
XX  
DT 29-JUL-2004 (first entry)  
XX  
DE Klebsiella pneumoniae polypeptide seqid 13283.  
XX

KW Recombinant expression vector; transcription regulatory element;  
 KW Klebsiella pneumoniae protein; antibacterial; Vaccine.  
 XX  
 OS Klebsiella pneumoniae.  
 XX  
 PN US6610836-B1.  
 XX  
 PD 26-AUG-2003.  
 XX  
 PF 27-JAN-2000; 2000US-00489039.  
 XX  
 PR 29-JAN-1999; 99US-0117747P.  
 XX  
 PA (GENO-) GENOME THERAPEUTICS CORP.  
 XX  
 PI Breton GL, Osborne M;  
 XX  
 DR WPI; 2003-895346/82.  
 DR N-PSDB; ABD00337.  
 XX  
 PT New nucleic acid encoding a Klebsiella pneumoniae polypeptide, useful for  
 PT preparing a vaccine composition against Klebsiella pneumoniae.  
 XX  
 PS Disclosure; SEQ ID NO 13283; 932pp; English.  
 XX  
 CC The invention describes a new isolated nucleic acid encoding a Klebsiella  
 CC pneumoniae polypeptide. Also described are: a recombinant expression  
 CC vector comprising the nucleic acid, operably linked to a transcription  
 CC regulatory element; and a cell comprising the recombinant expression  
 CC vector. The nucleic acid is useful for preparing a vaccine composition  
 CC against Klebsiella pneumoniae. This is the amino acid sequence of a  
 CC Klebsiella pneumoniae polypeptide of the invention  
 XX  
 SQ Sequence 851 AA;

Query Match 53.8%; Score 43; DB 7; Length 851;  
 Best Local Similarity 63.6%; Pred. No. 2.5e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNFFGAGLDHQ 14  
 ||:| ||: ||  
 Db 572 GNWFSAGMTHQ 582

RESULT 3  
 ABU18793

ID ABU18793 standard; protein; 513 AA.  
 XX  
 AC ABU18793;  
 XX  
 DT 19-JUN-2003 (first entry)  
 XX  
 DE Protein encoded by Prokaryotic essential gene #4320.  
 XX  
 KW Antisense; prokaryotic essential gene; cell proliferation; drug design.  
 XX  
 OS Bacillus anthracis.



XX  
PN WO200277183-A2.  
XX  
PD 03-OCT-2002.  
XX  
PF 21-MAR-2002; 2002WO-US009107.  
XX  
PR 21-MAR-2001; 2001US-00815242.  
PR 06-SEP-2001; 2001US-00948993.  
PR 25-OCT-2001; 2001US-0342923P.  
PR 08-FEB-2002; 2002US-00072851.  
PR 06-MAR-2002; 2002US-0362699P.  
XX  
PA (ELIT-) ELITRA PHARM INC.  
XX  
PI Wang L, Zamudio C, Malone C, Haselbeck R, Ohlsen KL, Zyskind JW;  
PI Wall D, Trawick JD, Carr GJ, Yamamoto R, Forsyth RA, Xu HH;  
XX  
DR WPI; 2003-029926/02.  
DR N-PSDB; ACA22663.  
XX  
PT New antisense nucleic acids, useful for identifying proteins or screening  
PT for homologous nucleic acids required for cellular proliferation to  
PT isolate candidate molecules for rational drug discovery programs.  
XX  
PS Claim 25; SEQ ID NO 46717; 1766pp; English.  
XX  
CC The invention relates to an isolated nucleic acid comprising any one of  
CC the 6213 antisense sequences given in the specification where expression  
CC of the nucleic acid inhibits proliferation of a cell. Also included are:  
CC (1) a vector comprising a promoter operably linked to the nucleic acid  
CC encoding a polypeptide whose expression is inhibited by the antisense  
CC nucleic acid; (2) a host cell containing the vector; (3) an isolated  
CC polypeptide or its fragment whose expression is inhibited by the  
CC antisense nucleic acid; (4) an antibody capable of specifically binding  
CC the polypeptide; (5) producing the polypeptide; (6) inhibiting cellular  
CC proliferation or the activity of a gene in an operon required for  
CC proliferation; (7) identifying a compound that influences the activity of  
CC the gene product or that has an activity against a biological pathway  
CC required for proliferation, or that inhibits cellular proliferation; (8)  
CC identifying a gene required for cellular proliferation or the biological  
CC pathway in which a proliferation-required gene or its gene product lies  
CC or a gene on which the test compound that inhibits proliferation of an  
CC organism acts; (9) manufacturing an antibiotic; (10) profiling a  
CC compound's activity; (11) a culture comprising strains in which the gene  
CC product is overexpressed or underexpressed; (12) determining the extent  
CC to which each of the strains is present in a culture or collection of  
CC strains; or (13) identifying the target of a compound that inhibits the  
CC proliferation of an organism. The antisense nucleic acids are useful for  
CC identifying proteins or screening for homologous nucleic acids required  
CC for cellular proliferation to isolate candidate molecules for rational  
CC drug discovery programs, or for screening homologous nucleic acids  
CC required for proliferation in cells other than *S. aureus*, *S. typhimurium*,  
CC *K. pneumoniae* or *P. aeruginosa*. The present sequence is encoded by one of  
CC the target prokaryotic essential genes. Note: The sequence data for this  
CC patent did not form part of the printed specification, but was obtained  
CC in electronic format directly from WIPO at

CC ftp.wipo.int/pub/published\_pct\_sequences

XX

SQ Sequence 513 AA;

Query Match 52.5%; Score 42; DB 6; Length 513;

Best Local Similarity 63.6%; Pred. No. 2.1e+02;

Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNFFGAGLDHQ 14

|:| | | | |:

Db 367 GSFFGLGLHHK 377

#### RESULT 4

AAW23689

ID AAW23689 standard; protein; 171 AA.

XX

AC AAW23689;

XX

DT 23-MAR-1998 (first entry)

XX

DE Potato polyphenol oxidase GPOT10.

XX

KW Polyphenol oxidase; PPO; browning; fruit; vegetable;  
KW genomic DNA amplification.

XX

OS Solanum tuberosum.

XX

PN WO9729193-A1.

XX

PD 14-AUG-1997.

XX

PF 24-JAN-1997; 97WO-AU000041.

XX

PR 05-FEB-1996; 96AU-00007856.

PR 16-SEP-1996; 96AU-00002361.

XX

PA (CSIR ) COMMONWEALTH SCI & IND RES ORG.

XX

PI Robinson SP;

XX

DR WPI; 1997-415348/38.

DR N-PSDB; AAT78400.

XX

PT Preparing nucleic acid encoding polyphenol oxidase - by genomic DNA  
PT amplification, useful to control browning reactions in fruit and  
PT vegetables.

XX

PS Claim 40; Fig 26; 53pp; English.

XX

CC A method has been developed for preparing a nucleic acid sequence  
CC encoding polyphenol oxidase (PPO), or a fragment or derivative. The  
CC method comprises amplifying genomic DNA isolated from plant tissue with  
CC sense and antisense primers corresponding to conserved PPO gene regions.  
CC The present sequence represents a specifically claimed polyphenol  
CC oxidase. Sense nucleic acid sequences can be used to increase or, by co-  
CC suppression, decrease PPO activity in plants, while antisense nucleic

CC acid sequences reduce activity. Control of PPO activity allows browning  
CC reactions in fruit and vegetables to be controlled, while avoiding the  
CC need for chemicals, e.g. sulphur dioxide. Many PPO genes lack introns,  
CC and can therefore be amplified directly from genomic DNA, eliminating the  
CC need for separation of RNA and synthesis of cDNA. Also, only small  
CC samples are needed and fragment size can be predicted, allowing bands of  
CC appropriate size to be selected for cloning

XX

SQ Sequence 171 AA;

Query Match 51.2%; Score 41; DB 2; Length 171;  
Best Local Similarity 77.8%; Pred. No. 95;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 4 GNFFGAGLD 12  
|||: |||  
Db 121 GNFYSAGLD 129

#### RESULT 5

AAAY70004

ID AAY70004 standard; protein; 215 AA.

XX

AC AAY70004;

XX

DT 31-MAY-2000 (first entry)

XX

DE Casein kinase II beta subunit.

XX

KW Plasmid pRAB-84-69; pRSB-14; recombinant; beta-casein;  
KW casein kinase II alpha subunit; casein kinase II beta subunit;  
KW kanamycin resistance marker; iminopeptidase; genetic stability;  
KW pharmaceutical; nutritional composition; vaccine formulation.

XX

OS Unidentified.

XX

PN WO200008174-A1.

XX

PD 17-FEB-2000.

XX

PF 06-AUG-1999; 99WO-US017873.

XX

PR 07-AUG-1998; 98US-00131028.

XX

PA (ABBO ) ABBOTT LAB.

XX

PI Mukerji P, Lemmel SA, Leonard AE, Chaudhary S;

XX

DR WPI; 2000-205721/18.

DR N-PSDB; AAZ50910, AAZ50911.

XX

PT Recombinant construct useful for producing human milk protein, edible  
PT plant protein, antibody, antigen or hormone, comprises nucleotide  
PT sequences expressing beta-casein protein.

XX

PS Disclosure; Fig 7; 73pp; English.

XX

CC The patent discloses a method of producing human milk protein, edible  
CC plant protein, antibody or an antigen in a host cell. It involves  
CC transforming host cells with a vector comprising the gene of interest  
CC linked to a promoter and nucleotide sequences encoding subunits of a  
CC kinase, resistance marker and a peptidase. This method is useful for  
CC improving the genetic stability of a plasmid-containing cell during  
CC fermentation. Proteins produced may be used in pharmaceutical or  
CC nutritional compositions and in vaccine formulations. The present  
CC sequence is that of casein kinase II beta subunit encoded by plasmid  
CC constructs pRAB-84-69 and pRSB-14. These constructs also express  
CC recombinant human beta-casein, casein kinase II alpha subunit, bacterial  
CC kanamycin resistance marker and iminopeptidase  
XX  
SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 3; Length 215;  
Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
::||:||||  
Db 153 HTDGAYFGTGPPH 165

RESULT 6

ABB66949

ID ABB66949 standard; protein; 215 AA.

XX

AC ABB66949;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 27639.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL11052.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell

PT interactions.  
XX  
PS Disclosure; SEQ ID NO 27639; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 4; Length 215;  
Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGFP 165

#### RESULT 7

ABB66950

ID ABB66950 standard; protein; 215 AA.

XX

AC ABB66950;

XX

DT 26-MAR-2002 (first entry)

XX

DE Drosophila melanogaster polypeptide SEQ ID NO 27642.

XX

KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.

XX

OS Drosophila melanogaster.

XX

PN WO200171042-A2.

XX

PD 27-SEP-2001.

XX

PF 23-MAR-2001; 2001WO-US009231.

XX

PR 23-MAR-2000; 2000US-0191637P.

PR 11-JUL-2000; 2000US-00614150.

XX

PA (PEKE ) PE CORP NY.

XX

PI Venter JC, Adams M, Li PWD, Myers EW;

XX

DR WPI; 2001-656860/75.

DR N-PSDB; ABL11053.

XX

PT New isolated nucleic acid detection reagent for detecting 1000 or more  
PT genes from Drosophila and for elucidating cell signaling and cell-cell  
PT interactions.  
XX  
PS Disclosure; SEQ ID NO 27642; 21pp + Sequence Listing; English.  
XX  
CC The invention relates to an isolated nucleic acid detection reagent  
CC capable of detecting 1000 or more genes from Drosophila. The invention is  
CC useful in developmental biology and in elucidating cell signalling and  
CC cell-cell interactions in higher eukaryotes for the development of  
CC insecticides, therapeutics and pharmaceutical drugs. The invention  
CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
CC ABB72072). The sequence data for this patent did not form part of the  
CC printed specification, but was obtained in electronic format directly  
CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
XX  
SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 4; Length 215;  
Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||: || | |  
Db 153 HTDGAYFGTGFPH 165

RESULT 3  
ABB70309

ID ABB70309 standard; protein; 215 AA.  
XX  
AC ABB70309;  
XX  
DT 26-MAR-2002 (first entry)  
XX  
DE Drosophila melanogaster polypeptide SEQ ID NO 37719.  
XX  
KW Drosophila; developmental biology; cell signalling; insecticide;  
KW pharmaceutical.  
XX  
OS Drosophila melanogaster.  
XX  
PN WO200171042-A2.  
XX  
PD 27-SEP-2001.  
XX  
PF 23-MAR-2001; 2001WO-US009231.  
XX  
PR 23-MAR-2000; 2000US-0191637P.  
PR 11-JUL-2000; 2000US-00614150.  
XX  
PA (PEKE ) PE CORP NY.  
XX  
PI Venter JC, Adams M, Li PWD, Myers EW;  
XX  
DR WPI; 2001-656860/75.

DR N-PSDB; ABL14412.  
 XX  
 PT New isolated nucleic acid detection reagent for detecting 1000 or more  
 PT genes from Drosophila and for elucidating cell signaling and cell-cell  
 PT interactions.  
 XX  
 PS Disclosure; SEQ ID NO 37719; 21pp + Sequence Listing; English.  
 XX  
 CC The invention relates to an isolated nucleic acid detection reagent  
 CC capable of detecting 1000 or more genes from Drosophila. The invention is  
 CC useful in developmental biology and in elucidating cell signalling and  
 CC cell-cell interactions in higher eukaryotes for the development of  
 CC insecticides, therapeutics and pharmaceutical drugs. The invention  
 CC discloses genomic DNA sequences (ABL16176-ABL30511), expressed DNA  
 CC sequences (ABL01840-ABL16175) and the encoded proteins (ABB57737-  
 CC ABB72072). The sequence data for this patent did not form part of the  
 CC printed specification, but was obtained in electronic format directly  
 CC from WIPO at ftp.wipo.int/pub/published\_pct\_sequences  
 XX  
 SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 4; Length 215;  
 Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
 ::||:|||  
 Db 153 HTDGAYFGTGFP 165

RESULT 9  
 ABG79668

ID ABG79668 standard; protein; 215 AA.  
 XX  
 AC ABG79668;  
 XX  
 DT 15-NOV-2002 (first entry)  
 XX  
 DE Human casein kinase 2-beta protein.  
 XX  
 KW Casein kinase2-beta; antisense gene therapy; cytostatic; enzyme;  
 KW antidiabetic; antiinflammatory; diabetes; cancer; tumour; breast cancer;  
 KW hyperproliferative disorder; prostate cancer; liver cancer; human.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200262954-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 31-JAN-2002; 2002WO-US003159.  
 XX  
 PR 08-FEB-2001; 2001US-00780175.  
 XX  
 PA (ISIS-) ISIS PHARM INC.  
 XX  
 PI McKay R, Freier SM, Wyatt JR;

XX  
 DR WPI; 2002-643409/69.  
 DR N-PSDB; ABS65048, ABG65062.  
 XX  
 PT New antisense oligonucleotides targeted to nucleic acid encoding Casein  
 PT kinase 2-beta, useful in diagnostic and research applications, or for  
 PT treating a disease or condition associated with the expression of Casein  
 PT kinase 2-beta.  
 XX  
 PS Example 15; Page 106-107; 142pp; English.  
 XX  
 CC The invention relates to a compound that is 8 - 50 nucleobases in length  
 CC targeted to a nucleic acid molecule encoding Casein kinase 2-beta, and  
 CC which specifically hybridises with and inhibits the expression of Casein  
 CC kinase 2-beta, or which specifically hybridises with an 8-nucleobase  
 CC portion of an active site on a nucleic acid molecule encoding Casein  
 CC kinase 2-beta. Also included are: (1) a composition comprising the  
 CC compound, and a carrier or diluent; (2) inhibiting the expression of  
 CC Casein kinase 2-beta in cells or tissues by contacting the cells or  
 CC tissues with the compound so that the expression of Casein kinase 2-beta  
 CC is inhibited; and (3) treating an animal having a disease or condition  
 CC associated with Casein kinase 2-beta by administering to the animal the  
 CC new compound so that the expression of Casein kinase 2-beta is inhibited.  
 CC The antisense compounds are useful for modulating the expression of  
 CC Casein kinase 2-beta and for treating diseases or conditions associated  
 CC with expression of Casein kinase 2-beta, e.g. diabetes or  
 CC hyperproliferative disorders, particularly cancer, such as breast cancer,  
 CC prostate cancer, or liver cancer. The antisense compounds are also useful  
 CC for diagnostics, therapeutics, prophylaxis, e.g. to prevent or delay  
 CC infection, inflammation or tumour formation, as research reagents and  
 CC kits, and in distinguishing between functions of various members of a  
 CC biological pathway. The present sequence is the casein kinase 2-beta  
 CC protein sequence  
 XX  
 SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 5; Length 215;  
 Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
 ::|| :|| | |  
 Db 153 HTDGAYFGTGFP 165

# RESULT 10

ABG79669

ID ABG79669 standard; protein; 215 AA.

XX

AC ABG79669;

XX

DT 15-NOV-2002 (first entry)

XX

DE Mouse casein kinase 2-beta protein.

XX

KW Casein kinase2-beta; antisense gene therapy; cytostatic; enzyme;

KW antidiabetic; antiinflammatory; diabetes; cancer; tumour; breast cancer;



KW hyperproliferative disorder; prostate cancer; liver cancer; mouse.  
 XX  
 OS Mus musculus.  
 XX  
 PN WO200262954-A2.  
 XX  
 PD 15-AUG-2002.  
 XX  
 PF 31-JAN-2002; 2002WO-US003159.  
 XX  
 PR 08-FEB-2001; 2001US-00780175.  
 XX  
 PA (ISIS-) ISIS PHARM INC.  
 XX  
 PI McKay R, Freier SM, Wyatt JR;  
 XX  
 DR WPI; 2002-643409/69.  
 DR N-PSDB; ABS65055, ABG65141.  
 XX  
 PT New antisense oligonucleotides targeted to nucleic acid encoding Casein  
 PT kinase 2-beta, useful in diagnostic and research applications, or for  
 PT treating a disease or condition associated with the expression of Casein  
 PT kinase 2-beta.  
 XX  
 PS Example 16; Page 127-129; 142pp; English.  
 XX  
 CC The invention relates to a compound that is 8 - 50 nucleobases in length  
 CC targeted to a nucleic acid molecule encoding Casein kinase 2-beta, and  
 CC which specifically hybridises with and inhibits the expression of Casein  
 CC kinase 2-beta, or which specifically hybridises with an 8-nucleobase  
 CC portion of an active site on a nucleic acid molecule encoding Casein  
 CC kinase 2-beta. Also included are: (1) a composition comprising the  
 CC compound, and a carrier or diluent; (2) inhibiting the expression of  
 CC Casein kinase 2-beta in cells or tissues by contacting the cells or  
 CC tissues with the compound so that the expression of Casein kinase 2-beta  
 CC is inhibited; and (3) treating an animal having a disease or condition  
 CC associated with Casein kinase 2-beta by administering to the animal the  
 CC new compound so that the expression of Casein kinase 2-beta is inhibited.  
 CC The antisense compounds are useful for modulating the expression of  
 CC Casein kinase 2-beta and for treating diseases or conditions associated  
 CC with expression of Casein kinase 2-beta, e.g. diabetes or  
 CC hyperproliferative disorders, particularly cancer, such as breast cancer,  
 CC prostate cancer, or liver cancer. The antisense compounds are also useful  
 CC for diagnostics, therapeutics, prophylaxis, e.g. to prevent or delay  
 CC infection, inflammation or tumour formation, as research reagents and  
 CC kits, and in distinguishing between functions of various members of a  
 CC biological pathway. The present sequence is the casein kinase 2-beta  
 CC protein sequence  
 XX  
 SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 5; Length 215;  
 Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
 ::||:|||

Db 153 HTDGAYFGTGFP 165

RESULT 11

ADB79767

ID ADB79767 standard; protein; 215 AA.

XX

AC ADB79767;

XX

DT 04-DEC-2003 (first entry)

XX

DE Rat Casein kinase II beta subunit, SEQ ID 7.

XX

KW Analgesic; pain; streptozocin-induced diabetes; rat.

XX

OS Rattus norvegicus.

XX

PN EP1279744-A2.

XX

PD 29-JAN-2003.

XX

PF 26-JUL-2002; 2002EP-00255249.

XX

PR 27-JUL-2001; 2001GB-00018354.

PR 07-FEB-2002; 2002GB-00002910.

XX

PA (WARN ) WARNER LAMBERT CO.

XX

PI Brooksbank RA, Dixon AK, Lee K, Pinnock RD;

XX

DR WPI; 2003-395407/38.

DR N-PSDB; ADB79768.

XX

PT Use of isolated gene sequences and encoded polypeptides that are  
PT upregulated in the spinal cord in response to streptozocin-induced  
PT diabetes for screening compounds for the treatment of pain, or for  
PT diagnosing pain.

XX

PS Claim 1; Page 49-50; 334pp; English.

XX

CC The present invention relates to nucleotide sequences which are useful in  
CC the screening of compounds for the treatment of pain, or for the  
CC diagnosis of pain. The nucleotide sequences are up-regulated in the  
CC spinal cord in response to streptozocin-induced diabetes. The present  
CC sequence was used to illustrate the invention.

XX

SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 7; Length 215;

Best Local Similarity 46.2%; Pred. No. 1.2e+02;

Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13

::||:|||

Db 153 HTDGAYFGTGFP 165

RESULT 12

ADC53769

ID ADC53769 standard; protein; 215 AA.

XX

AC ADC53769;

XX

DT 18-DEC-2003 (first entry)

XX

DE Casein kinase-2 beta protein, SEQ ID No 3.

XX

KW enzyme; casein kinase-2; p53DINP1; phosphorylate; Ser46;

KW cancer suppression protein; p53; cancer; cytostatic.

XX

OS Unidentified.

XX

PN JP2003093056-A.

XX

PD 02-APR-2003.

XX

PF 26-SEP-2001; 2001JP-00292953.

XX

PR 26-SEP-2001; 2001JP-00292953.

XX

PA (KAGA-) KAGAKU GIJUTSU SHINKO JIGYODAN.

PA (KOKU-) KOKURITSU GAN CENT SOCHO.

XX

DR WPI; 2003-590918/56.

XX

PT Novel enzyme useful as cancer therapeutic agent and as screening agent

PT for identifying anti-cancer agents, comprises casein kinase-2 and

PT p53DINP1.

XX

PS Claim 2; SEQ ID NO 3; 12pp; Japanese.

XX

CC The invention relates to a novel enzyme which consists of casein kinase-2  
 CC and p53DINP1 that phosphorylates Ser46 of the cancer suppression protein,  
 CC p53. The novel enzyme is useful as a cancer therapeutic agent and as a  
 CC screening agent for chemical compounds which activate p53 or which  
 CC inhibit activity of the enzyme. The enzyme has cytostatic activity. This  
 CC sequence represents a casein kinase-2 enzyme protein of the invention.

XX

SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 7; Length 215;

Best Local Similarity 46.2%; Pred. No. 1.2e+02;

Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13

::||:| | |

Db 153 HTDGAYFGTGFPH 165

RESULT 13

ADE57086

ID ADE57086 standard; protein; 215 AA.

XX

AC ADE57086;

XX  
 DT 29-JAN-2004 (first entry)  
 XX  
 DE Human Protein P13862, SEQ ID NO 2946.  
 XX  
 KW Human; pain; neuronal tissue; gene therapy;  
 KW spinal segmental nerve injury; chronic constriction injury; CCI;  
 KW spared nerve injury; SNI; Chung.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO2003016475-A2.  
 XX  
 PD 27-FEB-2003.  
 XX  
 PF 14-AUG-2002; 2002WO-US025765.  
 XX  
 PR 14-AUG-2001; 2001US-0312147P.  
 PR 01-NOV-2001; 2001US-0346382P.  
 PR 26-NOV-2001; 2001US-0333347P.  
 XX  
 PA (GEHO ) GEN HOSPITAL CORP.  
 PA (FARB ) BAYER AG.  
 XX  
 PI Woolf C, D'urso D, Befort K, Costigan M;  
 XX  
 DR WPI; 2003-268312/26.  
 DR GENBANK; P13862.  
 XX  
 PT New composition comprising two or more isolated polypeptides, useful for  
 PT preparing a medicament for treating pain in an animal.  
 XX  
 PS Claim 1; Page; 1017pp; English.  
 XX  
 CC The invention discloses a composition comprising two or more isolated rat  
 CC or human polynucleotides or a polynucleotide which represents a fragment,  
 CC derivative or allelic variation of the nucleic acid sequence. Also  
 CC claimed are a vector comprising the novel polynucleotide, a host cell  
 CC comprising the vector, a method for identifying a nucleotide sequence  
 CC which is differentially regulated in an animal subjected to pain and a  
 CC kit to perform the method, an array, a method for identifying an agent  
 CC that increases or decreases the expression of the polynucleotide sequence  
 CC that is differentially expressed in neuronal tissue of a first animal  
 CC subjected to pain, a method for identifying a compound which regulates  
 CC the expression of a polynucleotide sequence which is differentially  
 CC expressed in an animal subjected to pain, a method for identifying a  
 CC compound that regulates the activity of one or more of the  
 CC polynucleotides, a method for producing a pharmaceutical composition, a  
 CC method for identifying a compound or small molecule that regulates the  
 CC activity in an animal of one or more of the polypeptides given in the  
 CC specification, a method for identifying a compound useful in treating  
 CC pain and a pharmaceutical composition comprising the one or more  
 CC polypeptides or their antibodies. The polynucleotide or the compound that  
 CC modulates its activity is useful for preparing a medicament for treating  
 CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
 CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
 CC therapy). The sequence presented is a human protein (shown in Table 2 of

CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 7; Length 215;  
Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGFP 165

#### RESULT 14

ADE57088

ID ADE57088 standard; protein; 215 AA.

XX

AC ADE57088;

XX

DT 29-JAN-2004 (first entry)

XX

DE Rat Protein P13862, SEQ ID NO 2948.

XX

KW Rat; pain; neuronal tissue; gene therapy; spinal segmental nerve injury;  
KW chronic constriction injury; CCI; spared nerve injury; SNI; Chung.

XX

OS Rattus norvegicus.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX

PR 14-AUG-2001; 2001US-0312147P.

PR 01-NOV-2001; 2001US-0346382P.

PR 26-NOV-2001; 2001US-0333347P.

XX

PA (GEHO ) GEN HOSPITAL CORP.

PA (FARB ) BAYER AG.

XX

PI Woolf C, D'urso D, Befort K, Costigan M;

XX

DR WPI; 2003-268312/26.

DR GENBANK; P13862.

XX

PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.

XX

PS Claim 1; Page; 1017pp; English.

XX

CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also

CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a rat protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.

XX

SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 7; Length 215;  
Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGFP 165

#### RESULT 15

ADD47777

ID ADD47777 standard; protein; 215 AA.

XX

AC ADD47777;

XX

DT 29-JAN-2004 (first entry)

XX

DE Human Protein P13862, SEQ ID NO 13473.

XX

KW Human; pain; neuronal tissue; gene therapy;

KW spinal segmental nerve injury; chronic constriction injury; CCI;

KW spared nerve injury; SNI; Chung.

XX

OS Homo sapiens.

XX

PN WO2003016475-A2.

XX

PD 27-FEB-2003.

XX

PF 14-AUG-2002; 2002WO-US025765.

XX  
PR 14-AUG-2001; 2001US-0312147P.  
PR 01-NOV-2001; 2001US-0346382P.  
PR 26-NOV-2001; 2001US-0333347P.  
XX  
PA (GEHO ) GEN HOSPITAL CORP.  
PA (FARB ) BAYER AG.  
XX  
PI Woolf C, D'urso D, Befort K, Costigan M;  
XX  
DR WPI; 2003-268312/26.  
DR GENBANK; P13862.  
XX  
PT New composition comprising two or more isolated polypeptides, useful for  
PT preparing a medicament for treating pain in an animal.  
XX  
PS Claim 1; Page; 1017pp; English.  
XX  
CC The invention discloses a composition comprising two or more isolated rat  
CC or human polynucleotides or a polynucleotide which represents a fragment,  
CC derivative or allelic variation of the nucleic acid sequence. Also  
CC claimed are a vector comprising the novel polynucleotide, a host cell  
CC comprising the vector, a method for identifying a nucleotide sequence  
CC which is differentially regulated in an animal subjected to pain and a  
CC kit to perform the method, an array, a method for identifying an agent  
CC that increases or decreases the expression of the polynucleotide sequence  
CC that is differentially expressed in neuronal tissue of a first animal  
CC subjected to pain, a method for identifying a compound which regulates  
CC the expression of a polynucleotide sequence which is differentially  
CC expressed in an animal subjected to pain, a method for identifying a  
CC compound that regulates the activity of one or more of the  
CC polynucleotides, a method for producing a pharmaceutical composition, a  
CC method for identifying a compound or small molecule that regulates the  
CC activity in an animal of one or more of the polypeptides given in the  
CC specification, a method for identifying a compound useful in treating  
CC pain and a pharmaceutical composition comprising the one or more  
CC polypeptides or their antibodies. The polynucleotide or the compound that  
CC modulates its activity is useful for preparing a medicament for treating  
CC pain (e.g. spinal segmental nerve injury (Chung), chronic constriction  
CC injury (CCI) and spared nerve injury (SNI)) in an animal (e.g. gene  
CC therapy). The sequence presented is a human protein (shown in Table 2 of  
CC the specification) which is differentially expressed during pain. Note:  
CC The sequence data for this patent did not form part of the printed  
CC specification, but was obtained in electronic form directly from WIPO at  
CC ftp.wipo.int/pub/published\_pct\_sequences.  
XX  
SQ Sequence 215 AA;

Query Match 51.2%; Score 41; DB 7; Length 215;  
Best Local Similarity 46.2%; Pred. No. 1.2e+02;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGFP 165

Search completed: January 31, 2005, 13:17:05  
Job time : 94.9545 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 13:08:40 ; Search time 24.8182 Seconds  
(without alignments)  
37.410 Million cell updates/sec

Title: US-10-067-620-8  
Perfect score: 80  
Sequence: 1 YSDGNFFGAGLDHQ 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 478139 seqs, 66318000 residues

Total number of hits satisfying chosen parameters: 478139

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/1/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result | Score | Match | Length | DB | ID                   | Description       |
|--------|-------|-------|--------|----|----------------------|-------------------|
| No.    |       | %     |        |    |                      |                   |
| 1      | 43    | 53.8  | 851    | 4  | US-09-489-039A-13283 | Sequence 13283, A |
| 2      | 41    | 51.2  | 171    | 3  | US-09-129-030-56     | Sequence 56, Appl |
| 3      | 41    | 51.2  | 215    | 3  | US-09-131-028A-3     | Sequence 3, Appli |
| 4      | 41    | 51.2  | 215    | 3  | US-09-131-028A-13    | Sequence 13, Appl |
| 5      | 41    | 51.2  | 215    | 4  | US-09-538-092-923    | Sequence 923, App |
| 6      | 41    | 51.2  | 223    | 4  | US-09-443-067-12     | Sequence 12, Appl |
| 7      | 41    | 51.2  | 227    | 4  | US-09-443-067-14     | Sequence 14, Appl |
| 8      | 41    | 51.2  | 379    | 3  | US-09-028-934-36     | Sequence 36, Appl |
| 9      | 41    | 51.2  | 438    | 4  | US-08-482-934A-12    | Sequence 12, Appl |



|    |      |      |      |   |                      |                   |
|----|------|------|------|---|----------------------|-------------------|
| 10 | 41   | 51.2 | 596  | 3 | US-08-481-190-8      | Sequence 8, Appli |
| 11 | 41   | 51.2 | 596  | 5 | PCT-US93-00869-8     | Sequence 8, Appli |
| 12 | 40   | 50.0 | 321  | 4 | US-09-489-039A-9579  | Sequence 9579, Ap |
| 13 | 40   | 50.0 | 429  | 4 | US-09-328-352-4643   | Sequence 4643, Ap |
| 14 | 40   | 50.0 | 494  | 3 | US-09-198-956-2      | Sequence 2, Appli |
| 15 | 40   | 50.0 | 494  | 4 | US-09-670-141-2      | Sequence 2, Appli |
| 16 | 40   | 50.0 | 741  | 4 | US-09-489-039A-7854  | Sequence 7854, Ap |
| 17 | 40   | 50.0 | 810  | 3 | US-09-540-824-25     | Sequence 25, Appl |
| 18 | 39.5 | 49.4 | 836  | 4 | US-09-489-039A-7701  | Sequence 7701, Ap |
| 19 | 39.5 | 49.4 | 1041 | 4 | US-09-644-827B-9     | Sequence 9, Appli |
| 20 | 39.5 | 49.4 | 1084 | 4 | US-09-637-145-3      | Sequence 3, Appli |
| 21 | 39.5 | 49.4 | 1084 | 4 | US-09-538-092-1222   | Sequence 1222, Ap |
| 22 | 39   | 48.8 | 241  | 4 | US-09-252-991A-21735 | Sequence 21735, A |
| 23 | 39   | 48.8 | 400  | 4 | US-09-270-767-31765  | Sequence 31765, A |
| 24 | 39   | 48.8 | 400  | 4 | US-09-270-767-46982  | Sequence 46982, A |
| 25 | 39   | 48.8 | 582  | 3 | US-09-450-072-80     | Sequence 80, Appl |
| 26 | 39   | 48.8 | 582  | 4 | US-09-351-348-80     | Sequence 80, Appl |
| 27 | 39   | 48.8 | 1001 | 4 | US-09-252-991A-26291 | Sequence 26291, A |
| 28 | 38   | 47.5 | 102  | 4 | US-09-252-991A-21025 | Sequence 21025, A |
| 29 | 38   | 47.5 | 337  | 2 | US-08-861-464-12     | Sequence 12, Appl |
| 30 | 38   | 47.5 | 337  | 2 | US-08-396-001-12     | Sequence 12, Appl |
| 31 | 38   | 47.5 | 337  | 3 | US-09-323-433A-12    | Sequence 12, Appl |
| 32 | 38   | 47.5 | 337  | 4 | US-09-826-752-12     | Sequence 12, Appl |
| 33 | 38   | 47.5 | 353  | 4 | US-09-252-991A-27528 | Sequence 27528, A |
| 34 | 38   | 47.5 | 452  | 4 | US-09-252-991A-20884 | Sequence 20884, A |
| 35 | 38   | 47.5 | 675  | 4 | US-09-252-991A-27026 | Sequence 27026, A |
| 36 | 38   | 47.5 | 766  | 2 | US-08-553-436A-8     | Sequence 8, Appli |
| 37 | 38   | 47.5 | 822  | 4 | US-09-248-796A-14728 | Sequence 14728, A |
| 38 | 37.5 | 46.9 | 421  | 4 | US-09-252-991A-32440 | Sequence 32440, A |
| 39 | 37   | 46.2 | 60   | 4 | US-09-489-039A-13332 | Sequence 13332, A |
| 40 | 37   | 46.2 | 401  | 4 | US-09-328-352-5448   | Sequence 5448, Ap |
| 41 | 37   | 46.2 | 429  | 4 | US-09-252-991A-32753 | Sequence 32753, A |
| 42 | 37   | 46.2 | 474  | 4 | US-09-489-039A-10432 | Sequence 10432, A |
| 43 | 37   | 46.2 | 518  | 4 | US-09-252-991A-23491 | Sequence 23491, A |
| 44 | 37   | 46.2 | 521  | 4 | US-09-252-991A-27868 | Sequence 27868, A |
| 45 | 37   | 46.2 | 720  | 4 | US-09-394-272-14     | Sequence 14, Appl |

#### ALIGNMENTS

##### RESULT 1

US-09-489-039A-13283

; Sequence 13283, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO KLEBSIELLA

; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS

; FILE REFERENCE: 2709.2004001

; CURRENT APPLICATION NUMBER: US/09/489,039A

; CURRENT FILING DATE: 2000-01-27

; PRIOR APPLICATION NUMBER: US 60/117,747

; PRIOR FILING DATE: 1999-01-29

; NUMBER OF SEQ ID NOS: 14342

; SEQ ID NO 13283

; LENGTH: 851  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-13283

Query Match 53.8%; Score 43; DB 4; Length 851;  
Best Local Similarity 63.6%; Pred. No. 54;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GNFFGAGLDHQ 14  
||: ||: ||  
Db 572 GNWFSAGMTHQ 582

RESULT 2

US-09-129-030-56  
; Sequence 56, Application US/09129030A  
; Patent No. 6242221  
; GENERAL INFORMATION:  
; APPLICANT: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH ORGANISATION  
; TITLE OF INVENTION: GENOMIC PPO CLONES  
; FILE REFERENCE: 57072-PCT-US  
; CURRENT APPLICATION NUMBER: US/09/129,030A  
; CURRENT FILING DATE: 1998-08-04  
; EARLIER APPLICATION NUMBER: AU PN7856  
; EARLIER FILING DATE: 1996-02-05  
; EARLIER APPLICATION NUMBER: AU PO2361  
; EARLIER FILING DATE: 1996-09-16  
; EARLIER APPLICATION NUMBER: PCT/AU97/00041  
; EARLIER FILING DATE: 1997-01-24  
; NUMBER OF SEQ ID NOS: 66  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 56  
; LENGTH: 171  
; TYPE: PRT  
; ORGANISM: POTATO  
US-09-129-030-56

Query Match 51.2%; Score 41; DB 3; Length 171;  
Best Local Similarity 77.8%; Pred. No. 20;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GNFFGAGLD 12  
|||: |||  
Db 121 GNFYSAGLD 129

RESULT 3

US-09-131-028A-3  
; Sequence 3, Application US/09131028A  
; Patent No. 6287866  
; GENERAL INFORMATION:  
; APPLICANT: Abbott Laboratories  
; APPLICANT: Mukerji, Pradip  
; APPLICANT: Lemmel, Steven A.  
; APPLICANT: Leonard, Amanda Eun-Yeong  
; APPLICANT: Chaudhary, Sunita

; TITLE OF INVENTION: BETA-CASEIN EXPRESSING CONSTRUCTS  
; FILE REFERENCE: 6004.US.P1  
; CURRENT APPLICATION NUMBER: US/09/131,028A  
; CURRENT FILING DATE: 1998-08-07  
; PRIOR APPLICATION NUMBER: US 08/064,440  
; PRIOR FILING DATE: 1993-05-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 3  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-131-028A-3

Query Match 51.2%; Score 41; DB 3; Length 215;  
Best Local Similarity 46.2%; Pred. No. 26;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGPH 165

RESULT 4

US-09-131-028A-13  
; Sequence 13, Application US/09131028A  
; Patent No. 6287866  
; GENERAL INFORMATION:  
; APPLICANT: Abbott Laboratories  
; APPLICANT: Mukerji, Pradip  
; APPLICANT: Lemmel, Steven A.  
; APPLICANT: Leonard, Amanda Eun-Yeong  
; APPLICANT: Chaudhary, Sunita  
; TITLE OF INVENTION: BETA-CASEIN EXPRESSING CONSTRUCTS  
; FILE REFERENCE: 6004.US.P1  
; CURRENT APPLICATION NUMBER: US/09/131,028A  
; CURRENT FILING DATE: 1998-08-07  
; PRIOR APPLICATION NUMBER: US 08/064,440  
; PRIOR FILING DATE: 1993-05-21  
; NUMBER OF SEQ ID NOS: 22  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 13  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-131-028A-13

Query Match 51.2%; Score 41; DB 3; Length 215;  
Best Local Similarity 46.2%; Pred. No. 26;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGPH 165

RESULT 5

US-09-538-092-923

; Sequence 923, Application US/09538092  
; Patent No. 6753314  
; GENERAL INFORMATION:  
; APPLICANT: Giot, Loic  
; APPLICANT: Mansfield, Traci A.  
; TITLE OF INVENTION: Protein-Protein Complexes and Method of Using Same  
; FILE REFERENCE: 15966-542  
; CURRENT APPLICATION NUMBER: US/09/538,092  
; CURRENT FILING DATE: 2000-03-29  
; PRIOR APPLICATION NUMBER: 60/127,352  
; PRIOR FILING DATE: 1999-04-01  
; PRIOR APPLICATION NUMBER: 60/178,965  
; PRIOR FILING DATE: 2000-02-01  
; NUMBER OF SEQ ID NOS: 1387  
; SOFTWARE: CuraPatSeqFormatter Version 0.9  
; SEQ ID NO 923  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (0)...(0)  
; OTHER INFORMATION: Polypeptide Accession Number P13862  
US-09-538-092-923

Query.Match 51.2%; Score 41; DB 4; Length 215;  
Best Local Similarity 46.2%; Pred. No. 26;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGFP 165

RESULT 6

US-09-443-067-12

; Sequence 12, Application US/09443067  
; Patent No. 6627794  
; GENERAL INFORMATION:  
; APPLICANT: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH  
; APPLICANT: ORGANISATION  
; TITLE OF INVENTION: Polyphenol oxidase genes from banana, lettuce, tobacco  
and  
; TITLE OF INVENTION: pineapple  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/443,067  
; CURRENT FILING DATE: 1999-11-18  
; EARLIER APPLICATION NUMBER: US 08/976, 222  
; EARLIER FILING DATE: 1997-11-21  
; EARLIER APPLICATION NUMBER: PCT/AU98/00362  
; EARLIER FILING DATE: 1998-05-19  
; EARLIER APPLICATION NUMBER: AU PP3898  
; EARLIER FILING DATE: 1995-05-23  
; EARLIER APPLICATION NUMBER: AU PP6849  
; EARLIER FILING DATE: 1997-05-19  
; EARLIER APPLICATION NUMBER: AU PP5600

; EARLIER FILING DATE: 1995-09-26  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 12  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: tobacco  
US-09-443-067-12

Query Match 51.2%; Score 41; DB 4; Length 223;  
Best Local Similarity 77.8%; Pred. No. 27;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GNFFGAGLD 12  
|||: ||||  
Db 173 GNFYSAGLD 181

RESULT 7

US-09-443-067-14  
; Sequence 14, Application US/09443067  
; Patent No. 6627794  
; GENERAL INFORMATION:  
; APPLICANT: COMMONWEALTH SCIENTIFIC AND INDUSTRIAL RESEARCH  
; APPLICANT: ORGANISATION  
; TITLE OF INVENTION: Polyphenol oxidase genes from banana, lettuce, tobacco  
and  
; TITLE OF INVENTION: pineapple  
; FILE REFERENCE:  
; CURRENT APPLICATION NUMBER: US/09/443,067  
; CURRENT FILING DATE: 1999-11-18  
; EARLIER APPLICATION NUMBER: US 08/976, 222  
; EARLIER FILING DATE: 1997-11-21  
; EARLIER APPLICATION NUMBER: PCT/AU98/00362  
; EARLIER FILING DATE: 1998-05-19  
; EARLIER APPLICATION NUMBER: AU PP3898  
; EARLIER FILING DATE: 1995-05-23  
; EARLIER APPLICATION NUMBER: AU PP6849  
; EARLIER FILING DATE: 1997-05-19  
; EARLIER APPLICATION NUMBER: AU PP5600  
; EARLIER FILING DATE: 1995-09-26  
; NUMBER OF SEQ ID NOS: 49  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 14  
; LENGTH: 227  
; TYPE: PRT  
; ORGANISM: tobacco  
US-09-443-067-14

Query Match 51.2%; Score 41; DB 4; Length 227;  
Best Local Similarity 77.8%; Pred. No. 28;  
Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GNFFGAGLD 12  
|||: ||||  
Db 177 GNFYSAGLD 185

RESULT 8

US-09-028-934-36

; Sequence 36, Application US/09028934

; Patent No. 6117670

; GENERAL INFORMATION:

; APPLICANT: Ligon, James M.

; APPLICANT: Hill, Dwight S.

; APPLICANT: Lam, Steven T.

; APPLICANT: Hammer, Philip E.

; APPLICANT: van Pee, Karl-Heinz

; APPLICANT: Kirner, Sabine

; APPLICANT: Young, Thomas R.

; TITLE OF INVENTION: Pyrrolnitrin Biosynthesis Genes and Uses

; TITLE OF INVENTION: Thereof

; NUMBER OF SEQUENCES: 37

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 6117670artis Corporation

; STREET: 3054 Cornwallis Road

; CITY: Research Triangle Park

; STATE: NC

; COUNTRY: USA

; ZIP: 27709

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.30

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/028,934

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/729,214

; FILING DATE: 09-OCT-1996

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: US 08/258,261

; FILING DATE: 08-JUN-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: Meigs, J. Timothy

; REGISTRATION NUMBER: 38,241

; REFERENCE/DOCKET NUMBER: CGC1506/CIP7

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 919-541-8587

; TELEFAX: 919-541-8689

; INFORMATION FOR SEQ ID NO: 36:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 379 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-09-028-934-36

Query Match 51.2%; Score 41; DB 3; Length 379;

Best Local Similarity 70.0%; Pred. No. 49;

Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 3 DGNFFGAGLD 12  
|| :|||:|  
Db 217 DGAWFGAGID 226

RESULT 9

US-08-482-934A-12

; Sequence 12, Application US/08482934A  
; Patent No. 6703542

; GENERAL INFORMATION:

; APPLICANT: Robinson, Simon P.

; APPLICANT: Dry, Ian B.

; TITLE OF INVENTION: Polyphenol Oxidase Genes

; NUMBER OF SEQUENCES: 27

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Cooper & Dunham LLP

; STREET: 1185 Avenue of the Americas

; CITY: New York

; STATE: New York

; COUNTRY: USA

; ZIP: 10036

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/482,934A

; FILING DATE: 07-JUN-1995

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/182,045

; FILING DATE: 14-FEB-1994

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: WO PCT/AU92/00356

; FILING DATE: 16-JUL-1992

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: AU PK7248

; FILING DATE: 17-JUL-1991

; ATTORNEY/AGENT INFORMATION:

; NAME: John P. White

; REGISTRATION NUMBER: 28,678

; REFERENCE/DOCKET NUMBER: 51461-Z/JPW/GJG

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-278-0400

; TELEFAX: 212-391-0526

; INFORMATION FOR SEQ ID NO: 12:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 438 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-482-934A-12

Query Match 51.2%; Score 41; DB 4; Length 438;

Best Local Similarity 77.8%; Pred. No. 57;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GNFFGAGLD 12  
|||: |||  
Db 153 GNFYSAGLD 161

RESULT 10

US-08-481-190-8

; Sequence 8, Application US/08481190

; Patent No. 6160204

; GENERAL INFORMATION:

; APPLICANT: John C. Steffens

; TITLE OF INVENTION: Polyphenol Oxidase cDNA

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Yahwak & Associates

; STREET: 25 Skytop Drive

; CITY: Trumbull

; STATE: Connecticut

; COUNTRY: USA

; ZIP: 06611

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; COMPUTER: Macintosh

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Microsoft Word 4.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/481,190

; FILING DATE:

; CLASSIFICATION:

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 203,533

; FILING DATE: 02-24-1994

; ATTORNEY/AGENT INFORMATION:

; NAME: George M. Yahwak

; REGISTRATION NUMBER: 26,824

; REFERENCE/DOCKET NUMBER: UA 816 CIP

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (203)268-1951

; TELEFAX: (203)268-1951

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 596 amino acids

; TYPE: amino acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

US-08-481-190-8

Query Match 51.2%; Score 41; DB 3; Length 596;

Best Local Similarity 77.8%; Pred. No. 80;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GNFFGAGLD 12  
|||: |||  
Db 356 GNFYSAGLD 364



RESULT 11

PCT-US93-00869-8

; Sequence 8, Application PC/TUS9300869

; GENERAL INFORMATION:

; APPLICANT: John C. Steffens

; TITLE OF INVENTION: Polyphenol Oxidase cDNAs: Cloning

; TITLE OF INVENTION: and Applications

; NUMBER OF SEQUENCES: 19

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Yahwak & Associates

; STREET: 25 Skytop Drive

; CITY: Trumbull

; STATE: Connecticut

; COUNTRY: USA

; ZIP: 06611

; COMPUTER READABLE FORM:

; MEDIUM TYPE: floppy disk

; COMPUTER: Macintosh

; OPERATING SYSTEM: MS-DOS

; SOFTWARE: Microsoft Word 4.0

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: PCT/US93/00869

; FILING DATE: 19930129

; CLASSIFICATION:

; ATTORNEY/AGENT INFORMATION:

; NAME: George M. Yahwak

; REGISTRATION NUMBER: 26,824

; REFERENCE/DOCKET NUMBER: CRF D-1057

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (203)268-1951

; TELEFAX: (203)268-1951

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 596 amino acids

; TYPE: AMINO ACID

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: peptide

PCT-US93-00869-8

Query Match 51.2%; Score 41; DB 5; Length 596;

Best Local Similarity 77.8%; Pred. No. 80;

Matches 7; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GNFFGAGLD 12

|||: |||

Db 356 GNFYSAGLD 364

RESULT 12

US-09-489-039A-9579

; Sequence 9579, Application US/09489039A

; Patent No. 6610836

; GENERAL INFORMATION:

; APPLICANT: Gary Breton et. al

; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
KLEBSIELLA  
; TITLE OF INVENTION: PNEUMONIAE FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.2004001  
; CURRENT APPLICATION NUMBER: US/09/489,039A  
; CURRENT FILING DATE: 2000-01-27  
; PRIOR APPLICATION NUMBER: US 60/117,747  
; PRIOR FILING DATE: 1999-01-29  
; NUMBER OF SEQ ID NOS: 14342  
; SEQ ID NO 9579  
; LENGTH: 321  
; TYPE: PRT  
; ORGANISM: Klebsiella pneumoniae  
US-09-489-039A-9579

Query Match 50.0%; Score 40; DB 4; Length 321;  
Best Local Similarity 100.0%; Pred. No. 60;  
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 7 FGAGLDH 13  
|||  
Db 203 FGAGLDH 209

RESULT 13

US-09-328-352-4643  
; Sequence 4643, Application US/09328352  
; Patent No. 6562958  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO  
ACINETOBACTER  
; TITLE OF INVENTION: BAUMANNII FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: GTC99-03PA  
; CURRENT APPLICATION NUMBER: US/09/328,352  
; CURRENT FILING DATE: 1999-06-04  
; NUMBER OF SEQ ID NOS: 8252  
; SEQ ID NO 4643  
; LENGTH: 429  
; TYPE: PRT  
; ORGANISM: Acinetobacter baumannii  
US-09-328-352-4643

Query Match 50.0%; Score 40; DB 4; Length 429;  
Best Local Similarity 53.8%; Pred. No. 82;  
Matches 7; Conservative 3; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDGNFFGAGLDHQ 14  
|| :|| |::||  
Db 1 SDMSFFALGVNHQ 13

RESULT 14

US-09-198-956-2  
; Sequence 2, Application US/09198956  
; Patent No. 6165769  
; GENERAL INFORMATION:

```
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/198,956
; CURRENT FILING DATE: 1998-11-24
; EARLIER APPLICATION NUMBER: 1344/97
; EARLIER FILING DATE: 1997-11-24
; EARLIER APPLICATION NUMBER: 60/067,240
; EARLIER FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 494
;   TYPE: PRT
;   ORGANISM: Bacillus licheniformis
US-09-198-956-2
```

```
Query Match          50.0%; Score 40; DB 3; Length 494;
Best Local Similarity 60.0%; Pred. No. 96;
Matches      6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;
```

```
Qy      1 YSDGNFFGAG 10
        ||: ||:| |
Db      410 YSEANYFGTG 419
```

# RESULT 15

```
US-09-670-141-2
; Sequence 2, Application US/09670141
; Patent No. 6429000
; GENERAL INFORMATION:
; APPLICANT: Andersen, Lene N.
; APPLICANT: Schulein, Martin
; APPLICANT: Lange, Niels Erik K.
; APPLICANT: Bjornvad, Mads E.
; APPLICANT: Schnorr, Kirk
; TITLE OF INVENTION: Pectin Degrading Enzymes From Bacillus
; TITLE OF INVENTION: Licheniformis
; FILE REFERENCE: 5377.200-US
; CURRENT APPLICATION NUMBER: US/09/670,141
; CURRENT FILING DATE: 2000-09-26
; PRIOR APPLICATION NUMBER: 09/198,956
; PRIOR FILING DATE: 1998-11-24
; PRIOR APPLICATION NUMBER: 1344/97
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/067,240
; PRIOR FILING DATE: 1997-12-02
; NUMBER OF SEQ ID NOS: 26
; SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 2
;   LENGTH: 494
;   TYPE: PRT
```

; ORGANISM: Bacillus licheniformis  
US-09-670-141-2

Query Match 50.0%; Score 40; DB 4; Length 494;  
Best Local Similarity 60.0%; Pred. No. 96;  
Matches 6; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAG 10  
||: ||: ||  
Db 410 YSEANYFGTG 419

Search completed: January 31, 2005, 13:25:10  
Job time : 25.8182 secs

GenCore version 5.1.6  
Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 13:22:56 ; Search time 79.8636 Seconds  
(without alignments)  
63.334 Million cell updates/sec

Title: US-10-067-620-8  
Perfect score: 80  
Sequence: 1 YSDGNFFGAGLDHQ 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1608061 seqs, 361289386 residues

Total number of hits satisfying chosen parameters: 1608061

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published\_Applications\_AA:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/PCT\_NEW\_PUB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US06\_NEW\_PUB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US06\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US07\_NEW\_PUB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/PCTUS\_PUBCOMB.pep:\*
- 7: /cgn2\_6/ptodata/1/pubpaa/US08\_NEW\_PUB.pep:\*
- 8: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 9: /cgn2\_6/ptodata/1/pubpaa/US09A\_PUBCOMB.pep:\*
- 10: /cgn2\_6/ptodata/1/pubpaa/US09B\_PUBCOMB.pep:\*
- 11: /cgn2\_6/ptodata/1/pubpaa/US09C\_PUBCOMB.pep:\*
- 12: /cgn2\_6/ptodata/1/pubpaa/US09\_NEW\_PUB.pep:\*
- 13: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*

14: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
 15: /cgn2\_6/ptodata/1/pubpaa/US10C\_PUBCOMB.pep:\*  
 16: /cgn2\_6/ptodata/1/pubpaa/US10D\_PUBCOMB.pep:\*  
 17: /cgn2\_6/ptodata/1/pubpaa/US10\_NEW\_PUB.pep:\*  
 18: /cgn2\_6/ptodata/1/pubpaa/US11\_NEW\_PUB.pep:\*  
 19: /cgn2\_6/ptodata/1/pubpaa/US60\_NEW\_PUB.pep:\*  
 20: /cgn2\_6/ptodata/1/pubpaa/US60\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

| Result No. | Score | %<br>Query Match | Length | DB | ID                   | Description       |
|------------|-------|------------------|--------|----|----------------------|-------------------|
| 1          | 80    | 100.0            | 14     | 14 | US-10-067-484-8      | Sequence 8, Appli |
| 2          | 80    | 100.0            | 14     | 14 | US-10-067-620-8      | Sequence 8, Appli |
| 3          | 46    | 57.5             | 1069   | 14 | US-10-369-493-3869   | Sequence 3869, Ap |
| 4          | 44    | 55.0             | 100    | 15 | US-10-424-599-150715 | Sequence 150715,  |
| 5          | 44    | 55.0             | 234    | 14 | US-10-369-493-5137   | Sequence 5137, Ap |
| 6          | 44    | 55.0             | 235    | 14 | US-10-369-493-5138   | Sequence 5138, Ap |
| 7          | 42    | 52.5             | 57     | 15 | US-10-424-599-154688 | Sequence 154688,  |
| 8          | 42    | 52.5             | 507    | 14 | US-10-369-493-16558  | Sequence 16558, A |
| 9          | 42    | 52.5             | 513    | 15 | US-10-282-122A-46717 | Sequence 46717, A |
| 10         | 41    | 51.2             | 77     | 17 | US-10-425-115-326322 | Sequence 326322,  |
| 11         | 41    | 51.2             | 106    | 17 | US-10-425-115-364048 | Sequence 364048,  |
| 12         | 41    | 51.2             | 215    | 14 | US-10-205-219-7      | Sequence 7, Appli |
| 13         | 41    | 51.2             | 223    | 9  | US-09-205-658-211    | Sequence 211, App |
| 14         | 41    | 51.2             | 223    | 10 | US-09-963-693-211    | Sequence 211, App |
| 15         | 41    | 51.2             | 269    | 15 | US-10-264-049-2966   | Sequence 2966, Ap |
| 16         | 41    | 51.2             | 456    | 14 | US-10-238-075-314    | Sequence 314, App |
| 17         | 40    | 50.0             | 99     | 17 | US-10-425-115-294323 | Sequence 294323,  |
| 18         | 40    | 50.0             | 116    | 17 | US-10-425-115-259071 | Sequence 259071,  |
| 19         | 40    | 50.0             | 254    | 14 | US-10-369-493-2525   | Sequence 2525, Ap |
| 20         | 40    | 50.0             | 352    | 14 | US-10-442-017-2      | Sequence 2, Appli |
| 21         | 40    | 50.0             | 352    | 14 | US-10-442-017-3      | Sequence 3, Appli |
| 22         | 40    | 50.0             | 419    | 14 | US-10-369-493-12586  | Sequence 12586, A |
| 23         | 40    | 50.0             | 544    | 15 | US-10-424-599-148458 | Sequence 148458,  |
| 24         | 40    | 50.0             | 558    | 15 | US-10-425-114-57793  | Sequence 57793, A |
| 25         | 40    | 50.0             | 709    | 15 | US-10-161-493-76     | Sequence 76, Appl |
| 26         | 40    | 50.0             | 743    | 15 | US-10-282-122A-69539 | Sequence 69539, A |
| 27         | 39.5  | 49.4             | 233    | 17 | US-10-425-115-310079 | Sequence 310079,  |
| 28         | 39.5  | 49.4             | 403    | 17 | US-10-425-115-310081 | Sequence 310081,  |
| 29         | 39.5  | 49.4             | 967    | 9  | US-09-817-913-7      | Sequence 7, Appli |
| 30         | 39.5  | 49.4             | 967    | 9  | US-09-817-538-7      | Sequence 7, Appli |
| 31         | 39.5  | 49.4             | 967    | 10 | US-09-563-728A-30    | Sequence 30, Appl |
| 32         | 39.5  | 49.4             | 967    | 17 | US-10-870-587-7      | Sequence 7, Appli |
| 33         | 39.5  | 49.4             | 1030   | 14 | US-10-115-482-36     | Sequence 36, Appl |
| 34         | 39.5  | 49.4             | 1041   | 17 | US-10-814-160-9      | Sequence 9, Appli |
| 35         | 39.5  | 49.4             | 1084   | 10 | US-09-800-187-2      | Sequence 2, Appli |
| 36         | 39.5  | 49.4             | 1084   | 14 | US-10-072-094-7      | Sequence 7, Appli |
| 37         | 39.5  | 49.4             | 1084   | 14 | US-10-173-539-12     | Sequence 12, Appl |
| 38         | 39.5  | 49.4             | 1084   | 14 | US-10-172-094-7      | Sequence 7, Appli |
| 39         | 39.5  | 49.4             | 1084   | 15 | US-10-360-534-4      | Sequence 4, Appli |
| 40         | 39    | 48.8             | 123    | 17 | US-10-425-115-239628 | Sequence 239628,  |

|    |    |      |     |    |                      |                   |
|----|----|------|-----|----|----------------------|-------------------|
| 41 | 39 | 48.8 | 149 | 17 | US-10-425-115-291688 | Sequence 291688,  |
| 42 | 39 | 48.8 | 157 | 15 | US-10-424-599-268231 | Sequence 268231,  |
| 43 | 39 | 48.8 | 277 | 16 | US-10-437-963-120880 | Sequence 120880,  |
| 44 | 39 | 48.8 | 290 | 15 | US-10-425-114-43130  | Sequence 43130, A |
| 45 | 39 | 48.8 | 377 | 16 | US-10-602-898A-6     | Sequence 6, Appli |

#### ALIGNMENTS

##### RESULT 1

US-10-067-484-8

```
; Sequence 8, Application US/10067484
; Publication No. US20030170763A1
; GENERAL INFORMATION:
; APPLICANT: Buchanan, Bob B.
; APPLICANT: del Val, Gregorio
; APPLICANT: Frick, Oscar L.
; TITLE OF INVENTION: RAGWEED ALLERGENS
; FILE REFERENCE: 416272000200
; CURRENT APPLICATION NUMBER: US/10/067,484
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 60/266,686
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 14
; TYPE: PRT
; ORGANISM: Ragweed
```

US-10-067-484-8

```
Query Match          100.0%; Score 80; DB 14; Length 14;
Best Local Similarity 100.0%; Pred. No. 1.2e-06;
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      1 YSDGNFFGAGLDHQ 14
        |||||
Db      1 YSDGNFFGAGLDHQ 14
```

##### RESULT 2

US-10-067-620-8

```
; Sequence 8, Application US/10067620
; Publication No. US20030180225A1
; GENERAL INFORMATION:
; APPLICANT: Buchanan, Bob B.
; APPLICANT: del Val, Gregorio
; APPLICANT: Frick, Oscar L.
; APPLICANT: Teuber, Suzanne S.
; TITLE OF INVENTION: WALNUT AND RYEGRASS ALLERGENS
; FILE REFERENCE: 416272003400
; CURRENT APPLICATION NUMBER: US/10/067,620
; CURRENT FILING DATE: 2002-02-04
; PRIOR APPLICATION NUMBER: US 60/266,686
; PRIOR FILING DATE: 2001-02-05
; NUMBER OF SEQ ID NOS: 11
```

; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 8  
; LENGTH: 14  
; TYPE: PRT  
; ORGANISM: Ragweed  
US-10-067-620-8

Query Match 100.0%; Score 80; DB 14; Length 14;  
Best Local Similarity 100.0%; Pred. No. 1.2e-06;  
Matches 14; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDHQ 14  
||| ||| ||| |||  
Db 1 YSDGNFFGAGLDHQ 14

RESULT 3

US-10-369-493-3869

; Sequence 3869, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 3869  
; LENGTH: 1069  
; TYPE: PRT  
; ORGANISM: Neurospora crassa  
; FEATURE:  
; NAME/KEY: unsure  
; LOCATION: (1)..(1069)  
; OTHER INFORMATION: unsure at all Xaa locations  
US-10-369-493-3869

Query Match 57.5%; Score 46; DB 14; Length 1069;  
Best Local Similarity 66.7%; Pred. No. 67;  
Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 3 DGNFFGAGLDHQ 14  
|| ||| ||| :  
Db 223 DGTFFGFGLDRE 234

RESULT 4

US-10-424-599-150715

; Sequence 150715, Application US/10424599

; Publication No. US20040031072A1  
; GENERAL INFORMATION:  
; APPLICANT: La Rosa Thomas J  
; APPLICANT: Kovalic David K  
; APPLICANT: Zhou Yihua  
; APPLICANT: Cao Yongwei  
; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated  
With  
; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement  
; FILE REFERENCE: 38-21(53223)B  
; CURRENT APPLICATION NUMBER: US/10/424,599  
; CURRENT FILING DATE: 2003-04-28  
; NUMBER OF SEQ ID NOS: 285684  
; SEQ ID NO 150715  
; LENGTH: 100  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_10711C.1.pep  
US-10-424-599-150715

Query Match 55.0%; Score 44; DB 15; Length 100;  
Best Local Similarity 58.3%; Pred. No. 12;  
Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 3 DGNFFGAGLDHQ 14  
: | | | | |  
Db 55 EGGFFGGGFHHQ 66

# RESULT 5

US-10-369-493-5137  
; Sequence 5137, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 5137  
; LENGTH: 234  
; TYPE: PRT  
; ORGANISM: Caenorhabditis elegans  
US-10-369-493-5137

Query Match 55.0%; Score 44; DB 14; Length 234;  
Best Local Similarity 46.2%; Pred. No. 30;



Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||::|| | |  
Db 152 HTDGSYFGTGFPH 164

RESULT 6

US-10-369-493-5138

; Sequence 5138, Application US/10369493

; Publication No. US20030233675A1

; GENERAL INFORMATION:

; APPLICANT: Cao, Yongwei

; APPLICANT: Hinkle, Gregory J.

; APPLICANT: Slater, Steven C.

; APPLICANT: Goldman, Barry S.

; APPLICANT: Chen, Xianfeng

; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION OF

; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES

; FILE REFERENCE: 38-10(52052)B

; CURRENT APPLICATION NUMBER: US/10/369,493

; CURRENT FILING DATE: 2003-02-28

; PRIOR APPLICATION NUMBER: US 60/360,039

; PRIOR FILING DATE: 2002-02-21

; NUMBER OF SEQ ID NOS: 47374

; SEQ ID NO 5138

; LENGTH: 235

; TYPE: PRT

; ORGANISM: Caenorhabditis elegans

US-10-369-493-5138

Query Match 55.0%; Score 44; DB 14; Length 235;

Best Local Similarity 46.2%; Pred. No. 30;

Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||::|| | |  
Db 153 HTDGSYFGTGFPH 165

RESULT 7

US-10-424-599-154688

; Sequence 154688, Application US/10424599

; Publication No. US20040031072A1

; GENERAL INFORMATION:

; APPLICANT: La Rosa Thomas J

; APPLICANT: Kovalic David K

; APPLICANT: Zhou Yihua

; APPLICANT: Cao Yongwei

; TITLE OF INVENTION: Soy Nucleic Acid Molecules and Other Molecules Associated With

; TITLE OF INVENTION: Plants and Uses Thereof for Plant Improvement

; FILE REFERENCE: 38-21(53223)B

; CURRENT APPLICATION NUMBER: US/10/424,599

; CURRENT FILING DATE: 2003-04-28

; NUMBER OF SEQ ID NOS: 285684

; SEQ ID NO 154688  
; LENGTH: 57  
; TYPE: PRT  
; ORGANISM: Glycine max  
; FEATURE:  
; OTHER INFORMATION: Clone ID: PAT\_MRT3847\_110704C.1.pep  
US-10-424-599-154688

Query Match 52.5%; Score 42; DB 15; Length 57;  
Best Local Similarity 50.0%; Pred. No. 14;  
Matches 7; Conservative 1; Mismatches 6; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDHQ 14  
| | | | | | | | | |  
Db 8 YCDGKICGTALDHE 21

RESULT 8

US-10-369-493-16558  
; Sequence 16558, Application US/10369493  
; Publication No. US20030233675A1  
; GENERAL INFORMATION:  
; APPLICANT: Cao, Yongwei  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Chen, Xianfeng  
; TITLE OF INVENTION: EXPRESSION OF MICROBIAL PROTEINS IN PLANTS FOR PRODUCTION  
OF  
; TITLE OF INVENTION: PLANTS WITH IMPROVED PROPERTIES  
; FILE REFERENCE: 38-10(52052)B  
; CURRENT APPLICATION NUMBER: US/10/369,493  
; CURRENT FILING DATE: 2003-02-28  
; PRIOR APPLICATION NUMBER: US 60/360,039  
; PRIOR FILING DATE: 2002-02-21  
; NUMBER OF SEQ ID NOS: 47374  
; SEQ ID NO 16558  
; LENGTH: 507  
; TYPE: PRT  
; ORGANISM: Bacillus thuringiensis  
US-10-369-493-16558

Query Match 52.5%; Score 42; DB 14; Length 507;  
Best Local Similarity 63.6%; Pred. No. 1.5e+02;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GNFFGAGLDHQ 14  
| : | | | | | | :  
Db 364 GSFFGLGLHHK 374

RESULT 9

US-10-282-122A-46717  
; Sequence 46717, Application US/10282122A  
; Publication No. US20040029129A1  
; GENERAL INFORMATION:  
; APPLICANT: Wang, Liangsu

```

; APPLICANT: Zamudio, Carlos
; APPLICANT: Malone, Cheryl
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari
; APPLICANT: Zyskind, Judith
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John
; APPLICANT: Carr, Grant
; APPLICANT: Yamamoto, Robert
; APPLICANT: Forsyth, R.
; APPLICANT: Xu, H.
; TITLE OF INVENTION: Identification of Essential Genes in Microorganisms
; FILE REFERENCE: ELITRA.034A
; CURRENT APPLICATION NUMBER: US/10/282,122A
; CURRENT FILING DATE: 2003-02-20
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/230,335
; PRIOR FILING DATE: 2000-09-06
; PRIOR APPLICATION NUMBER: 60/230,347
; PRIOR FILING DATE: 2000-09-09
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/267,636
; PRIOR FILING DATE: 2001-02-09
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; Remaining Prior Application data removed - See File Wrapper or PALM.
; NUMBER OF SEQ ID NOS: 78614
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 46717
; LENGTH: 513
; TYPE: PRT
; ORGANISM: Bacillus anthracis
US-10-282-122A-46717

```

```

Query Match          52.5%; Score 42; DB 15; Length 513;
Best Local Similarity 63.6%; Pred. No. 1.5e+02;
Matches      7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

```

```

Qy      4 GNFFGAGLDHQ 14
        |:||| || |:
Db     367 GSFFGLGLHHK 377

```

```

RESULT 10
US-10-425-115-326322
; Sequence 326322, Application US/10425115
; Publication No. US20040214272A1

```

```
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 326322
; LENGTH: 77
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_60677C.1.pep
US-10-425-115-326322
```

```
Query Match          51.2%; Score 41; DB 17; Length 77;
Best Local Similarity 80.0%; Pred. No. 29;
Matches      8; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
```

```
QY      2 SDGNFFGAGL 11
        | ||| |||
Db      5 STGNFFAAGL 14
```

# RESULT 11

```
US-10-425-115-364048
; Sequence 364048, Application US/10425115
; Publication No. US20040214272A1
; GENERAL INFORMATION:
; APPLICANT: La Rosa, Thomas J.
; APPLICANT: Kovalic, David K.
; APPLICANT: Zhou, Yihua
; APPLICANT: Cao, Yongwei
; TITLE OF INVENTION: Nucleic Acid Molecules and Other Molecules Associated
With
; TITLE OF INVENTION: Plants
; FILE REFERENCE: 38-21(53222)B
; CURRENT APPLICATION NUMBER: US/10/425,115
; CURRENT FILING DATE: 2003-04-28
; NUMBER OF SEQ ID NOS: 369326
; SEQ ID NO 364048
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Zea mays
; FEATURE:
; OTHER INFORMATION: Clone ID: MRT4577_95182C.1.pep
US-10-425-115-364048
```

```
Query Match          51.2%; Score 41; DB 17; Length 106;
Best Local Similarity 46.2%; Pred. No. 41;
Matches      6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;
```

Qy 1 YSDGNFFGAGLDH 13  
::|| :|| | |  
Db 44 HTDGAYFGTGFPH 56

RESULT 12

US-10-205-219-7

; Sequence 7, Application US/10205219  
; Publication No. US20030138803A1  
; GENERAL INFORMATION:  
; APPLICANT: Warner-Lambert Company  
; APPLICANT: Lee, Kevin  
; APPLICANT: Dixon, Alistair  
; APPLICANT: Brooksbank, Robert  
; APPLICANT: Pinnock, Robert  
; TITLE OF INVENTION: Identification and Use of Molecules Implicated in Pain  
; FILE REFERENCE: WL-A-018200  
; CURRENT APPLICATION NUMBER: US/10/205,219  
; CURRENT FILING DATE: 2002-07-24  
; PRIOR APPLICATION NUMBER: GB 0118354.0  
; PRIOR FILING DATE: 2001-07-27  
; NUMBER OF SEQ ID NOS: 197  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 7  
; LENGTH: 215  
; TYPE: PRT  
; ORGANISM: Rattus norvegicus  
; FEATURE:  
; OTHER INFORMATION: Casein kinase II beta subunit  
US-10-205-219-7

Query Match 51.2%; Score 41; DB 14; Length 215;  
Best Local Similarity 46.2%; Pred. No. 87;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::|| :|| | |  
Db 153 HTDGAYFGTGFPH 165

RESULT 13

US-09-205-658-211

; Sequence 211, Application US/09205658  
; Patent No. US20010029617A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruvkun, Gary  
; APPLICANT: Ogg, Scott  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
; FILE REFERENCE: 00786/351004  
; CURRENT APPLICATION NUMBER: US/09/205,658  
; CURRENT FILING DATE: 1998-12-03  
; EARLIER APPLICATION NUMBER: 08/857,076  
; EARLIER FILING DATE: 1997-05-15  
; EARLIER APPLICATION NUMBER: 08/888,534  
; EARLIER FILING DATE: 1997-07-07  
; EARLIER APPLICATION NUMBER: US98/10080

; EARLIER FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 328  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 211  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: Ascoris suum  
US-09-205-658-211

Query Match 51.2%; Score 41; DB 9; Length 223;  
Best Local Similarity 46.2%; Pred. No. 90;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDGNFFGAGLDHQ 14  
: || : | || : |  
Db 182 ADGEYFWEGLEHE 194

RESULT 14

US-09-963-693-211  
; Sequence 211, Application US/09963693  
; Publication No. US20030181364A1  
; GENERAL INFORMATION:  
; APPLICANT: Ruvkun, Gary  
; APPLICANT: Ogg, Scott  
; TITLE OF INVENTION: THERAPEUTIC AND DIAGNOSTIC TOOLS FOR  
; TITLE OF INVENTION: IMPAIRED GLUCOSE TOLERANCE CONDITIONS  
; FILE REFERENCE: 00786/351004  
; CURRENT APPLICATION NUMBER: US/09/963,693  
; CURRENT FILING DATE: 2001-09-25  
; PRIOR APPLICATION NUMBER: US/09/205,658  
; PRIOR FILING DATE: 1998-12-03  
; PRIOR APPLICATION NUMBER: 08/857,076  
; PRIOR FILING DATE: 1997-05-15  
; PRIOR APPLICATION NUMBER: 08/888,534  
; PRIOR FILING DATE: 1997-07-07  
; PRIOR APPLICATION NUMBER: US98/10080  
; PRIOR FILING DATE: 1998-05-15  
; NUMBER OF SEQ ID NOS: 328  
; SOFTWARE: FastSEQ for Windows Version 4.0  
; SEQ ID NO 211  
; LENGTH: 223  
; TYPE: PRT  
; ORGANISM: Ascoris suum  
US-09-963-693-211

Query Match 51.2%; Score 41; DB 10; Length 223;  
Best Local Similarity 46.2%; Pred. No. 90;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 2 SDGNFFGAGLDHQ 14  
: || : | || : |  
Db 182 ADGEYFWEGLEHE 194

RESULT 15

US-10-264-049-2966

; Sequence 2966, Application US/10264049  
 ; Publication No. US20040005579A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Birse et al.  
 ; TITLE OF INVENTION: Nucleic Acids, Proteins, and Antibodies  
 ; FILE REFERENCE: PA133P1  
 ; CURRENT APPLICATION NUMBER: US/10/264,049  
 ; CURRENT FILING DATE: 2002-10-04  
 ; PRIOR APPLICATION NUMBER: PCT/US01/18569  
 ; PRIOR FILING DATE: 2001-06-07  
 ; PRIOR APPLICATION NUMBER: US 60/209,467  
 ; PRIOR FILING DATE: 2000-06-07  
 ; NUMBER OF SEQ ID NOS: 4360  
 ; SOFTWARE: PatentIn Ver. 3.1  
 ; SEQ ID NO 2966  
 ; LENGTH: 269  
 ; TYPE: PRT  
 ; ORGANISM: Homo sapiens  
 ; FEATURE:  
 ; NAME/KEY: MISC\_FEATURE  
 ; LOCATION: (2)  
 ; OTHER INFORMATION: Xaa equals any of the twenty naturally occurring L-amino acids  
 US-10-264-049-2966

Query Match 51.2%; Score 41; DB 15; Length 269;  
 Best Local Similarity 46.2%; Pred. No. 1.1e+02;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
 ::||:|||  
 Db 207 HTDGAYFGTGFP 219

Search completed: January 31, 2005, 13:44:52  
 Job time : 81.8636 secs

GenCore version 5.1.6  
 Copyright (c) 1993 - 2005 Compugen Ltd.

OM protein - protein search, using sw model

Run on: January 31, 2005, 13:07:55 ; Search time 18.4545 Seconds  
 (without alignments)  
 72.992 Million cell updates/sec

Title: US-10-067-620-8  
 Perfect score: 80  
 Sequence: 1 YSDGNFFGAGLDHQ 14

Scoring table: BLOSUM62  
 Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_79:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | % Query |        | DB | ID     | Description        |
|------------|-------|---------|--------|----|--------|--------------------|
|            |       | Match   | Length |    |        |                    |
| 1          | 44    | 55.0    | 234    | 2  | B87852 | protein kin-10 [im |
| 2          | 44    | 55.0    | 235    | 2  | T24317 | casein kinase II ( |
| 3          | 44    | 55.0    | 1556   | 2  | S76781 | glutamate synthase |
| 4          | 43    | 53.8    | 841    | 2  | A90669 | probable enzyme [i |
| 5          | 43    | 53.8    | 841    | 2  | D85519 | probable enzyme ya |
| 6          | 43    | 53.8    | 841    | 2  | C64755 | yagX protein - Esc |
| 7          | 42    | 52.5    | 233    | 2  | G86350 | protein F8K7.16 [i |
| 8          | 42    | 52.5    | 508    | 2  | T22836 | hypothetical prote |
| 9          | 42    | 52.5    | 543    | 2  | I40545 | oligopeptide ABC t |
| 10         | 42    | 52.5    | 717    | 2  | T35219 | probable membrane  |
| 11         | 41    | 51.2    | 178    | 2  | T00644 | hypothetical prote |
| 12         | 41    | 51.2    | 196    | 2  | S14725 | casein kinase II ( |
| 13         | 41    | 51.2    | 209    | 2  | A25828 | casein kinase II ( |
| 14         | 41    | 51.2    | 215    | 2  | JC7269 | protein kinase (EC |
| 15         | 41    | 51.2    | 215    | 2  | C38611 | casein kinase II ( |
| 16         | 41    | 51.2    | 215    | 2  | A39459 | casein kinase II ( |
| 17         | 41    | 51.2    | 215    | 2  | JN0556 | casein kinase II ( |
| 18         | 41    | 51.2    | 215    | 2  | S20405 | casein kinase II ( |
| 19         | 41    | 51.2    | 215    | 2  | S14724 | casein kinase II ( |
| 20         | 41    | 51.2    | 231    | 2  | AI0191 | probable exported  |
| 21         | 41    | 51.2    | 464    | 2  | B86079 | probable glycopori |
| 22         | 41    | 51.2    | 464    | 2  | C91232 | probable glycoprot |
| 23         | 41    | 51.2    | 583    | 2  | S30930 | catechol oxidase ( |
| 24         | 41    | 51.2    | 585    | 1  | S33544 | catechol oxidase ( |
| 25         | 41    | 51.2    | 587    | 1  | S33543 | catechol oxidase ( |
| 26         | 41    | 51.2    | 588    | 2  | S30929 | catechol oxidase ( |
| 27         | 41    | 51.2    | 592    | 2  | T03682 | catechol oxidase ( |
| 28         | 41    | 51.2    | 596    | 1  | S33540 | catechol oxidase ( |
| 29         | 41    | 51.2    | 599    | 2  | T07097 | catechol oxidase ( |
| 30         | 41    | 51.2    | 626    | 2  | S33541 | catechol oxidase ( |
| 31         | 41    | 51.2    | 630    | 1  | S33539 | catechol oxidase ( |
| 32         | 40    | 50.0    | 254    | 2  | T40159 | casein kinase ii,  |
| 33         | 40    | 50.0    | 810    | 2  | S67696 | probable membrane  |
| 34         | 40    | 50.0    | 2116   | 2  | T49818 | glutamate synthase |



|    |    |      |     |   |        |                    |
|----|----|------|-----|---|--------|--------------------|
| 35 | 39 | 48.8 | 196 | 2 | S76230 | hypothetical prote |
| 36 | 39 | 48.8 | 213 | 2 | C83243 | acyl carrier prote |
| 37 | 39 | 48.8 | 231 | 2 | E97740 | phosphatidylserine |
| 38 | 39 | 48.8 | 231 | 2 | E71678 | hypothetical prote |
| 39 | 39 | 48.8 | 330 | 2 | E81998 | pseudouridylate sy |
| 40 | 39 | 48.8 | 330 | 2 | H81225 | ribosomal large ch |
| 41 | 39 | 48.8 | 377 | 2 | T04089 | GTP-binding protei |
| 42 | 39 | 48.8 | 377 | 2 | T07376 | G-protein beta cha |
| 43 | 39 | 48.8 | 488 | 2 | T38481 | probable pre-mRNA  |
| 44 | 39 | 48.8 | 570 | 2 | S51404 | hypothetical prote |
| 45 | 39 | 48.8 | 895 | 2 | AI3483 | aconitate hydratas |

## ALIGNMENTS

### RESULT 1

B87852

protein kin-10 [imported] - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 10-May-2001 #sequence\_revision 10-May-2001 #text\_change 24-May-2001

C;Accession: B87852

R;anonymous, The C. elegans Sequencing Consortium.

Science 282, 2012-2018, 1998

A;Title: Genome sequence of the nematode C. elegans: a platform for investigating biology.

A;Reference number: A75000; MUID:99059613; PMID:9851916

A;Note: see websites genome.wustl.edu/gsc/C\_elegans/ and www.sanger.ac.uk/Projects/C\_elegans/ for a list of authors

A;Note: published errata appeared in Science 283, 35, 1999; Science 283, 2103, 1999; and Science 285, 1493, 1999

A;Accession: B87852

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-234 <STO>

A;Cross-references: GB:chr\_I; PIDN:CAB00056.1; PID:g3879276; GSPDB:GN00019

C;Genetics:

A;Gene: kin-10

A;Map position: 1

C;Superfamily: human casein kinase II beta chain

Query Match 55.0%; Score 44; DB 2; Length 234;

Best Local Similarity 46.2%; Pred. No. 4.8;

Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13

::||::|| | |

Db 152 HTDGSYFGTGFPH 164

### RESULT 2

T24317

casein kinase II (EC 2.7.1.-) beta chain - Caenorhabditis elegans

C;Species: Caenorhabditis elegans

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T24317; T24320; A41036; B41036

R;Lennard, N.

submitted to the EMBL Data Library, July 1996  
 A;Reference number: Z19874  
 A;Accession: T24317  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-235 <WIL>  
 A;Cross-references: UNIPROT:P28548; EMBL:Z75713; PIDN:CAB00053.1; GSPDB:GN00019; CESP:T01G9.6b  
 A;Experimental source: clone T01G9  
 A;Accession: T24320  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-58,60-235 <WI2>  
 A;Cross-references: EMBL:Z75713; PIDN:CAB00056.1; GSPDB:GN00019; CESP:T01G9.6a  
 A;Experimental source: clone T01G9  
 R;Hu, E.; Rubin, C.S.  
 J. Biol. Chem. 266, 19796-19802, 1991  
 A;Title: Casein kinase II from *Caenorhabditis elegans*. Cloning, characterization, and developmental regulation of the gene encoding the beta subunit.  
 A;Reference number: A41036; MUID:92011787; PMID:1918084  
 A;Accession: A41036  
 A;Molecule type: DNA  
 A;Residues: 1-58,60-141,'M',143-235 <HUA>  
 A;Cross-references: GB:M73827; NID:g156245; PIDN:AAA27983.1; PID:g156246  
 A;Accession: B41036  
 A;Molecule type: mRNA  
 A;Residues: 1-58,60-141,'M',143-235 <HU2>  
 A;Cross-references: GB:M73827; NID:g156245; PIDN:AAA27983.1; PID:g156246  
 C;Genetics:  
 A;Gene: CESP:T01G9.6b; CESP:T01G9.6a  
 A;Map position: 1  
 A;Introns: 6/3; 59/1; 123/1; 186/2  
 C;Superfamily: human casein kinase II beta chain  
 C;Keywords: autophosphorylation; phosphoprotein; phosphotransferase; serine/threonine-specific protein kinase

Query Match 55.0%; Score 44; DB 2; Length 235;  
 Best Local Similarity 46.2%; Pred. No. 4.8;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
 ::||::|| | |  
 Db 153 HTDGSYFGTGFPH 165

# RESULT 3

S76781  
 glutamate synthase (ferredoxin) (EC 1.4.7.1) precursor - *Synechocystis* sp.  
 (strain PCC 6803)  
 C;Species: *Synechocystis* sp.  
 A;Variety: PCC 6803  
 C;Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 12-Jul-2004  
 C;Accession: S76781  
 R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.; Hirose, M.; Sugiura, M.; Sasamoto, S.; Kimura, T.; Hosouchi, T.;

Matsuno, A.; Muraki, A.; Nakazaki, N.; Naruo, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda, M.; Tabata, S.  
 DNA Res. 3, 109-136, 1996  
 A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis sp. PCC6803. II. Sequence determination of the entire genome and assignment of potential protein-coding regions.  
 A;Reference number: S74322; MUID:97061201; PMID:8905231  
 A;Accession: S76781  
 A;Status: nucleic acid sequence not shown; translation not shown  
 A;Molecule type: DNA  
 A;Residues: 1-1556 <KAN>  
 A;Cross-references: UNIPROT:P55038; EMBL:D90916; GB:AB001339; NID:g1653715; PIDN:BAA18693.1; PID:g1653782  
 A;Experimental source: PCC 6803  
 A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
 C;Superfamily: glutamate synthase, large subunit  
 C;Keywords: 3Fe-4S; metalloprotein; oxidoreductase  
 F;1-36/Domain: propeptide #status predicted <PRO>  
 F;37-1556/Product: glutamate synthase #status predicted <MAT>  
 F;37/Active site: Cys #status predicted  
 F;1173,1179,1184/Binding site: 3Fe-4S cluster (Cys) (covalent) #status predicted

Query Match 55.0%; Score 44; DB 2; Length 1556;  
 Best Local Similarity 66.7%; Pred. No. 37;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLD 12  
 :||| ||||  
 Db 370 FSDGKIVGAGLD 381

#### RESULT 4

A90669

probable enzyme [imported] - Escherichia coli (strain O157:H7, substrain RIMD 0509952)

C;Species: Escherichia coli

C;Date: 18-Jul-2001 #sequence\_revision 18-Jul-2001 #text\_change 09-Jul-2004

C;Accession: A90669

R;Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.; Ohtsubo, E.; Nakayama, K.; Murata, T.; Tanaka, M.; Tobe, T.; Iida, T.; Takami, H.; Honda, T.; Sasakawa, C.; Ogasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.

DNA Res. 8, 11-22, 2001

A;Title: Complete genome sequence of enterohemorrhagic Escherichia coli O157:H7 and genomic comparison with a laboratory strain K-12.

A;Reference number: A99629; MUID:21156231; PMID:11258796

A;Accession: A90669

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-841 <HAY>

A;Cross-references: UNIPROT:Q8X6I4; GB:BA000007; PIDN:BAB33744.1; PID:g13359778; GSPDB:GN00154

A;Experimental source: strain O157:H7, substrain RIMD 0509952

C;Genetics:

A;Gene: ECs0321

Query Match 53.8%; Score 43; DB 2; Length 841;  
Best Local Similarity 63.6%; Pred. No. 28;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GNFFGAGLDHQ 14  
||:| ||: ||  
Db 562 GNWFSAGMTHQ 572

RESULT 5

D85519

probable enzyme yagX [imported] - Escherichia coli (strain O157:H7, substrain EDL933)

C;Species: Escherichia coli

C;Date: 16-Feb-2001 #sequence\_revision 16-Feb-2001 #text\_change 09-Jul-2004

C;Accession: D85519

R;Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, G.F.; Evans, P.S.; Gregor, J.; Kirkpatrick, H.A.; Posfai, G.; Hackett, J.; Klink, S.; Boutin, A.; Shao, Y.; Miller, L.; Grotbeck, E.J.; Davis, N.W.; Lim, A.; Dimalanta, E.; Potamousis, K.; Apodaca, J.; Anantharaman, T.S.; Lin, J.; Yen, G.; Schwartz, D.C.; Welch, R.A.; Blattner, F.R.

Nature 409, 529-533, 2001

A;Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A;Reference number: A85480; MUID:21074935; PMID:11206551

A;Accession: D85519

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-841 <STO>

A;Cross-references: UNIPROT:Q8X6I4; GB:AE005174; NID:g12513077; PIDN:AAG54615.1; GSPDB:GN00145; UWGP:Z0358

A;Experimental source: strain O157:H7, substrain EDL933

C;Genetics:

A;Gene: yagX

Query Match 53.8%; Score 43; DB 2; Length 841;  
Best Local Similarity 63.6%; Pred. No. 28;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GNFFGAGLDHQ 14  
||:| ||: ||  
Db 562 GNWFSAGMTHQ 572

RESULT 6

C64755

yagX protein - Escherichia coli (strain K-12)

C;Species: Escherichia coli

C;Date: 12-Sep-1997 #sequence\_revision 17-Sep-1997 #text\_change 09-Jul-2004

C;Accession: C64755

R;Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; Collado-Vides, J.; Glasner, J.D.; Rode, C.K.; Mayhew, G.F.; Gregor, J.; Davis, N.W.; Kirkpatrick, H.A.; Goeden, M.A.; Rose, D.J.; Mau, B.; Shao, Y. Science 277, 1453-1462, 1997

A;Title: The complete genome sequence of Escherichia coli K-12.

A;Reference number: A64720; MUID:97426617; PMID:9278503

A;Accession: C64755

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA  
A;Residues: 1-841 <BLAT>  
A;Cross-references: UNIPROT:P77802; GB:AE000136; GB:U00096; NID:g2367103;  
PIDN:AAC73394.1; PID:g1786484; UWGP:b0291  
A;Experimental source: strain K-12, substrain MG1655  
C;Genetics:  
A;Gene: yagX

Query Match 53.8%; Score 43; DB 2; Length 841;  
Best Local Similarity 63.6%; Pred. No. 28;  
Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 4 GNFFGAGLDHQ 14  
||:| ||: ||  
Db 562 GNWFSAGMTHQ 572

#### RESULT 7

G86350

protein F8K7.16 [imported] - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 02-Mar-2001 #sequence\_revision 02-Mar-2001 #text\_change 09-Jul-2004

C;Accession: G86350

R;Theologis; A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.;  
Alonso, J.; Altaf, H.; Araujo, R.; Bowman, C.L.; Brooks, S.Y.; Buehler, E.;  
Chan, A.; Chao, Q.; Chen, H.; Cheuk, R.F.; Chin, C.W.; Chung, M.K.; Conn, L.;  
Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Dunn, P.; Etgu, P.;  
Feldblyum, T.V.; Feng, J.; Fong, B.; Fujii, C.Y.; Gill, J.E.; Goldsmith, A.D.;  
Haas, B.; Hansen, N.F.; Hughes, B.; Huizar, L.

Nature 408, 816-820, 2000

A;Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.;  
Kim, C.J.; Koo, H.L.; Kremenetskaia, I.; Kurtz, D.B.; Kwan, A.; Lam, B.; Langin-  
Hooper, S.; Lee, A.; Lee, J.M.; Lenz, C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu,  
S.X.; Liu, Z.A.; Luros, J.S.; Maiti, R.; Marziali, A.; Militscher, J.; Miranda,  
M.; Nguyen, M.; Nierman, W.C.; Osborne, B.I.; Pai, G.; Peterson, J.; Pham, P.K.;  
Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.

A;Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.;  
Tallon, L.J.; Tambunga, G.; Toriumi, M.J.; Town, C.D.; Utterback, T.; van Aken,  
S.; Vaysberg, M.; Vysotskaia, V.S.; Walker, M.; Wu, D.; Yu, G.; Fraser, C.M.;  
Venter, J.C.; Davis, R.W.

A;Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis..

A;Reference number: A86141; MUID:21016719; PMID:11130712

A;Accession: G86350

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-233 <STO>

A;Cross-references: UNIPROT:Q9XI04; GB:AE005172; NID:g5263325; PIDN:AAD41427.1;  
GSPDB:GN00141

C;Genetics:

A;Gene: F8K7.16

A;Map position: 1

Query Match 52.5%; Score 42; DB 2; Length 233;  
Best Local Similarity 66.7%; Pred. No. 10;  
Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLD 12

Db                   :| |||||: ||  
14 HSIGNFFGSPLD 25

RESULT 8

T22836

hypothetical protein F57B7.4 - *Caenorhabditis elegans*

C;Species: *Caenorhabditis elegans*

C;Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 09-Jul-2004

C;Accession: T22836

R;Lennard, N.

submitted to the EMBL Data Library, June 1996

A;Reference number: Z19623

A;Accession: T22836

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-508 <WIL>

A;Cross-references: UNIPROT:Q20930; EMBL:Z74037; PIDN:CAA98493.1; GSPDB:GN00023;  
CESP:F57B7.4

A;Experimental source: clone F57B7

C;Genetics:

A;Gene: CESP:F57B7.4

A;Map position: 5

A;Introns: 45/3; 137/2; 221/2; 256/2; 306/2; 409/3; 451/3

Query Match                   52.5%;   Score 42;   DB 2;   Length 508;  
Best Local Similarity   61.5%;   Pred. No. 24;  
Matches   8;   Conservative   1;   Mismatches   4;   Indels   0;   Gaps   0;

QY           1 YSDGNFFGAGLDH 13

      ||   ||||   ||:

Db           31 YSYSNFFGISLDN 43

RESULT 9

I40545

oligopeptide ABC transporter (oligopeptide-binding protein) appA - *Bacillus subtilis*

C;Species: *Bacillus subtilis*

C;Date: 12-Aug-1996 #sequence\_revision 12-Aug-1996 #text\_change 09-Jul-2004

C;Accession: I40545; C69586

R;Koide, A.; Hoch, J.A.

Mol. Microbiol. 13, 417-426, 1994

A;Title: Identification of a second oligopeptide transport system in *Bacillus subtilis* and determination of its role in sporulation.

A;Reference number: I40543; MUID:95089678; PMID:7997159

A;Accession: I40545

A;Status: preliminary; translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-543 <RES>

A;Cross-references: UNIPROT:P42061; EMBL:U20909; NID:g677942; PIDN:AAA62358.1;  
PID:g677945

R;Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertero, M.G.; Bessieres, P.; Bolotin, A.; Borchert, S.; Boriss, R.; Boursier, L.; Brans, A.; Braun, M.; Brignell, S.C.; Bron, S.; Brouillet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Choi, S.K.; Codani, J.J.; Connerton, I.F.; Cummings, N.J.; Daniel, R.A.; Denizot, F.; Devine, K.M.;

Duesterhoeft, A.; Ehrlich, S.D.; Emmerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari, E.

Nature 390, 249-256, 1997

A;Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Galleron, N.; Ghim, S.Y.; Glaser, P.; Goffeau, A.; Golightly, E.J.; Grandi, G.; Guiseppi, G.; Guy, B.J.; Haga, K.; Haiech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.; Itaya, M.; Jones, L.; Joris, B.; Karamata, D.; Kasahara, Y.; Klaerr-Blanchard, M.; Klein, C.; Kobayashi, Y.; Koetter, P.; Koningstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardinois, S.

A;Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel, C.; Medigue, C.; Medina, N.; Mellado, R.P.; Mizuno, M.; Moestl, D.; Nakai, S.; Noback, M.; Noone, D.; O'Reilly, M.; Ogawa, K.; Ogiwara, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelle, D.; Porwollik, S.; Prescott, A.M.; Presecan, E.; Pujic, P.; Purnelle, B.; Rapoport, G.; Rey, M.; Reynolds, S.; Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon, E.

A;Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Seror, S.J.; Serror, P.; Shin, B.S.; Soldo, B.; Sorokin, A.; Tacconi, E.; Takagi, T.; Takahashi, H.; Takemaru, K.; Takeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama, S.; Vandenbol, M.; Vannier, F.; Vassarotti, A.; Viari, A.; Wambutt, R.; Wedler, E.; Wedler, H.; Weitzenegger, T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.

A;Authors: Yoshikawa, H.F.; Zumstein, E.; Yoshikawa, H.; Danchin, A.

A;Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.

A;Reference number: A69580; MUID:98044033; PMID:9384377

A;Accession: C69586

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-543 <KUN>

A;Cross-references: GB:Z99110; GB:AL009126; NID:g2633472; PIDN:CAB12995.1; PID:g2633492

A;Experimental source: strain 168

C;Genetics:

A;Gene: appA

C;Superfamily: dipeptide transport protein

Query Match 52.5%; Score 42; DB 2; Length 543;

Best Local Similarity 58.3%; Pred. No. 26;

Matches 7; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLD 12

| | | | : | |

Db 166 YKDGNFYNNALD 177

RESULT 10

T35219

probable membrane protein - *Streptomyces coelicolor*

C;Species: *Streptomyces coelicolor*

C;Date: 05-Nov-1999 #sequence\_revision 05-Nov-1999 #text\_change 09-Jul-2004

C;Accession: T35219

R;Seeger, K.J.; Harris, D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.

submitted to the EMBL Data Library, September 1998

A;Reference number: Z21572

A;Accession: T35219  
 A;Status: preliminary; translated from GB/EMBL/DDBJ  
 A;Molecule type: DNA  
 A;Residues: 1-717 <SEE>  
 A;Cross-references: UNIPROT:O86709; EMBL:AL031515; PIDN:CAA20624.1;  
 GSPDB:GN00070; SCOEDB:SC5C7.12  
 A;Experimental source: strain A3(2)  
 C;Genetics:  
 A;Gene: SCOEDB:SC5C7.12  
 C;Superfamily: Streptomyces coelicolor probable membrane protein SC5C7.12

Query Match 52.5%; Score 42; DB 2; Length 717;  
 Best Local Similarity 61.5%; Pred. No. 35;  
 Matches 8; Conservative 0; Mismatches 5; Indels 0; Gaps 0;

Qy 2 SDGNFFGAGLDHQ 14  
 | | | | | | |  
 Db 279 SRGNLFGGGADEQ 291

# RESULT 11

T00644

hypothetical protein F3I6.7 - Arabidopsis thaliana

C;Species: Arabidopsis thaliana (mouse-ear cress)

C;Date: 01-Feb-1999 #sequence\_revision 01-Feb-1999 #text\_change 09-Jul-2004

C;Accession: T00644

R;Federspiel, N.A.; Palm, C.J.; Conway, A.B.; Kurtz, D.B.; Conway, A.R.; Au, M.;  
 Araujo, R.; Buehler, E.; Dewar, K.; Feng, J.; Kim, C.; Li, Y.; Oji, O.; Osborne,  
 B.I.; Shima, P.; Sun, H.; Toriumi, M.; Vysotskaia, V.S.; Yu, G.; Ecker, J.;  
 Theologis, A.; Davis, R.W.

submitted to the EMBL Data Library, February 1998

A;Reference number: Z14197

A;Accession: T00644

A;Status: translated from GB/EMBL/DDBJ

A;Molecule type: DNA

A;Residues: 1-178 <FED>

A;Cross-references: UNIPROT:O48681; EMBL:AC002396; NID:g2749918; PID:g2829866;

GSPDB:GN00059; ATSP:F3I6.7

C;Genetics:

A;Gene: ATSP:F3I6.7

A;Map position: 1

A;Introns: 50/3; 87/1

C;Superfamily: Arabidopsis thaliana hypothetical protein F3I6.7

Query Match 51.2%; Score 41; DB 2; Length 178;  
 Best Local Similarity 50.0%; Pred. No. 12;  
 Matches 7; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDHQ 14  
 | | | | | : : :  
 Db 71 YVDGNGFAAQMEHR 84

# RESULT 12

S14725

casein kinase II (EC 2.7.1.-) beta chain - pig (fragment)

C;Species: Sus scrofa domestica (domestic pig)



C;Date: 21-Nov-1993 #sequence\_revision 19-Jan-1996 #text\_change 11-Jan-2000  
 C;Accession: S14725; S14478  
 R;Boldyreff, B.; Piontek, K.; Schmidt-Spaniol, I.; Issinger, O.G.  
 Biochim. Biophys. Acta 1088, 439-441, 1991  
 A;Title: The beta subunit of casein kinase II: cloning of cDNAs from murine and porcine origin and expression of the porcine sequence as a fusion protein.  
 A;Reference number: S14724; MUID:91198153; PMID:2015307  
 A;Accession: S14725  
 A;Molecule type: mRNA  
 A;Residues: 1-196 <BOL>  
 A;Cross-references: EMBL:X56503; NID:g1932; PIDN:CAA39858.1; PID:g1933  
 C;Superfamily: human casein kinase II beta chain  
 C;Keywords: autophosphorylation; phosphoprotein; phosphotransferase; serine/threonine-specific protein kinase

Query Match 51.2%; Score 41; DB 2; Length 196;  
 Best Local Similarity 46.2%; Pred. No. 13;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
 ::||:|||  
 Db 134 HTDGAYFGTGFPH 146

#### RESULT 13

A25828  
 casein kinase II (EC 2.7.1.-) beta chain - bovine  
 C;Species: Bos primigenius taurus (cattle)  
 C;Date: 22-Jul-1987 #sequence\_revision 22-Jul-1987 #text\_change 10-Dec-1999  
 C;Accession: A25828  
 R;Takio, K.; Kuenzel, E.A.; Walsh, K.A.; Krebs, E.G.  
 Proc. Natl. Acad. Sci. U.S.A. 84, 4851-4855, 1987  
 A;Title: Amino acid sequence of the beta subunit of bovine lung casein kinase II.  
 A;Reference number: A25828; MUID:87260887; PMID:3299375  
 A;Accession: A25828  
 A;Molecule type: protein  
 A;Residues: 1-209 <TAK>  
 C;Superfamily: human casein kinase II beta chain  
 C;Keywords: phosphotransferase; serine/threonine-specific protein kinase

Query Match 51.2%; Score 41; DB 2; Length 209;  
 Best Local Similarity 46.2%; Pred. No. 14;  
 Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
 ::||:|||  
 Db 149 HTDGAYFGTGFPH 161

#### RESULT 14

JC7269  
 protein kinase (EC 2.7.1.37) CK2 beta chain - common carp  
 N;Alternate names: serine(threonine) protein kinase CK2 beta chain  
 C;Species: Cyprinus carpio (common carp)  
 C;Date: 18-Aug-2000 #sequence\_revision 18-Aug-2000 #text\_change 02-Sep-2000  
 C;Accession: JC7269

R;Vera, M.I.; Kausel, G.; Barrera, R.; Leal, S.; Figueroa, J.; Quezada, C.  
Biochem. Biophys. Res. Commun. 271, 735-740, 2000  
A;Title: Seasonal adaptation modulates the expression of the protein kinase CK2  
beta subunit gene in the carp.  
A;Reference number: JC7269  
A;Accession: JC7269  
A;Molecule type: mRNA  
A;Residues: 1-215 <VER>  
A;Cross-references: GB:AF133088  
A;Experimental source: strain male  
C;Superfamily: human casein kinase II beta chain  
C;Keywords: growth regulation; phosphotransferase

Query Match 51.2%; Score 41; DB 2; Length 215;  
Best Local Similarity 46.2%; Pred. No. 14;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGFPH 165

RESULT 15

C38611

casein kinase II (EC 2.7.1.-) beta chain - chicken

C;Species: Gallus gallus (chicken)

C;Date: 23-Aug-1991 #sequence\_revision 23-Aug-1991 #text\_change 09-Jul-2004

C;Accession: C38611

R;Maridor, G.; Park, W.; Krek, W.; Nigg, E.A.

J. Biol. Chem. 266, 2362-2368, 1991

A;Title: Casein kinase II. cDNA sequences, developmental expression, and tissue  
distribution of mRNAs for alpha, 'alpha', and beta subunits of the chicken  
enzyme.

A;Reference number: A38611; MUID:91115855; PMID:1989988

A;Accession: C38611

A;Status: preliminary

A;Molecule type: mRNA

A;Residues: 1-215 <MAR>

A;Cross-references: UNIPROT:P13862; GB:M59458; GB:J05738; NID:g211535;

PIDN:AAA48692.1; PID:g211536

C;Superfamily: human casein kinase II beta chain

C;Keywords: autophosphorylation; phosphoprotein; phosphotransferase;  
serine/threonine-specific protein kinase

Query Match 51.2%; Score 41; DB 2; Length 215;  
Best Local Similarity 46.2%; Pred. No. 14;  
Matches 6; Conservative 3; Mismatches 4; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
::||:|||  
Db 153 HTDGAYFGTGFPH 165

Search completed: January 31, 2005, 13:23:47  
Job time : 19.4545 secs

OM protein - protein search, using sw model

Run on: January 31, 2005, 12:56:50 ; Search time 106.591 Seconds  
(without alignments)  
75.572 Million cell updates/sec

Title: US-10-067-620-8  
Perfect score: 80  
Sequence: 1 YSDGNFFGAGLDHQ 14

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 1825181 seqs, 575374646 residues

Total number of hits satisfying chosen parameters: 1825181

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : UniProt\_02:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

| Result No. | Score | % Match | Query Length | DB ID         | Description        |
|------------|-------|---------|--------------|---------------|--------------------|
| 1          | 46    | 57.5    | 1208         | 2 Q7S470      | Q7s470 neurospora  |
| 2          | 45    | 56.2    | 1245         | 2 Q8IAN1      | Q8ian1 plasmodium  |
| 3          | 44.5  | 55.6    | 1562         | 2 Q883V9      | Q883v9 pseudomonas |
| 4          | 44    | 55.0    | 189          | 2 Q89ZY7      | Q89zy7 bacteroides |
| 5          | 44    | 55.0    | 234          | 1 KC2B_CAEEEL | P28548 caenorhabdi |
| 6          | 44    | 55.0    | 1524         | 2 Q7VA01      | Q7va01 prochloroco |
| 7          | 44    | 55.0    | 1556         | 1 GLTS_SYNY3  | P55038 synechocyst |
| 8          | 44    | 55.0    | 2863         | 2 Q983H6      | Q983h6 rhizobium l |
| 9          | 43    | 53.8    | 245          | 2 Q8IIW5      | Q8iiw5 plasmodium  |
| 10         | 43    | 53.8    | 504          | 1 CPK4_ONCMY  | O93297 oncorhynchu |
| 11         | 43    | 53.8    | 714          | 2 Q7UJE2      | Q7uje2 rhodopirell |
| 12         | 43    | 53.8    | 841          | 1 YAGX_ECOLI  | P77802 escherichia |
| 13         | 43    | 53.8    | 841          | 2 Q7AHC2      | Q7ahc2 escherichia |
| 14         | 43    | 53.8    | 841          | 2 Q8CWC1      | Q8cwc1 escherichia |
| 15         | 43    | 53.8    | 841          | 2 Q8X6I4      | Q8x6i4 escherichia |
| 16         | 42    | 52.5    | 233          | 2 Q9XI04      | Q9xi04 arabidopsis |
| 17         | 42    | 52.5    | 386          | 2 Q7VJZ0      | Q7vjz0 helicobacte |

|    |    |      |      |   |            |                    |
|----|----|------|------|---|------------|--------------------|
| 18 | 42 | 52.5 | 513  | 2 | Q6HG42     | Q6hg42 bacillus th |
| 19 | 42 | 52.5 | 513  | 2 | Q734J8     | Q734j8 bacillus ce |
| 20 | 42 | 52.5 | 513  | 2 | Q81B23     | Q81b23 bacillus ce |
| 21 | 42 | 52.5 | 513  | 2 | Q81MZ1     | Q81mz1 bacillus an |
| 22 | 42 | 52.5 | 513  | 2 | AAS42315   | Aas42315 bacillus  |
| 23 | 42 | 52.5 | 513  | 2 | AAT32538   | Aat32538 bacillus  |
| 24 | 42 | 52.5 | 520  | 2 | Q8X0T1     | Q8x0t1 neurospora  |
| 25 | 42 | 52.5 | 543  | 1 | APPA_BACSU | P42061 bacillus su |
| 26 | 42 | 52.5 | 717  | 2 | O86709     | O86709 streptomyce |
| 27 | 42 | 52.5 | 1521 | 2 | Q7UZY3     | Q7uzy3 prochloroco |
| 28 | 41 | 51.2 | 128  | 2 | Q8GYP2     | Q8gyp2 arabidopsis |
| 29 | 41 | 51.2 | 178  | 2 | O48681     | O48681 arabidopsis |
| 30 | 41 | 51.2 | 215  | 1 | KC2B_BRARE | Q91398 brachydanio |
| 31 | 41 | 51.2 | 215  | 1 | KC2B_DROME | P08182 drosophila  |
| 32 | 41 | 51.2 | 215  | 1 | KC2B_HUMAN | P13862 homo sapien |
| 33 | 41 | 51.2 | 215  | 1 | KC2B_XENLA | P28021 xenopus lae |
| 34 | 41 | 51.2 | 215  | 2 | Q967X2     | Q967x2 ciona intes |
| 35 | 41 | 51.2 | 215  | 2 | Q71U52     | Q71u52 cyprinus ca |
| 36 | 41 | 51.2 | 215  | 2 | Q7SZF8     | Q7szf8 fugu rubrip |
| 37 | 41 | 51.2 | 215  | 2 | Q6DEU1     | Q6deu1 xenopus tro |
| 38 | 41 | 51.2 | 215  | 2 | AAF03911   | Aaf03911 mus muscu |
| 39 | 41 | 51.2 | 215  | 2 | AAF66446   | Aaf66446 cyprinus  |
| 40 | 41 | 51.2 | 215  | 2 | AAM29452   | Aam29452 drosophil |
| 41 | 41 | 51.2 | 215  | 2 | AAM50092   | Aam50092 homo sapi |
| 42 | 41 | 51.2 | 215  | 2 | AAH03775   | Aah03775 mus muscu |
| 43 | 41 | 51.2 | 215  | 2 | AAF48093   | Aaf48093 drosophil |
| 44 | 41 | 51.2 | 215  | 2 | BAB22445   | Bab22445 mus muscu |
| 45 | 41 | 51.2 | 215  | 2 | BAB27147   | Bab27147 mus muscu |

# ALIGNMENTS

## RESULT 1

Q7S470

ID Q7S470 PRELIMINARY; PRT; 1208 AA.

AC Q7S470;

DT 01-MAR-2004 (TrEMBLrel. 26, Created)

DT 01-MAR-2004 (TrEMBLrel. 26, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Hypothetical protein.

GN Name=NCU02202.1;

OS Neurospora crassa.

OC Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;

OC Sordariomycetidae; Sordariales; Sordariaceae; Neurospora.

OX NCBI\_TaxID=5141;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=OR74A;

RA Galagan J.E., Calvo S.E., Borkovich K.A., Selker E.U., Read N.D.,

RA Jaffe D., FitzHugh W., Ma L.-J., Smirnov S., Purcell S., Rehman B.,

RA Elkins T., Engels R., Wang S., Nielsen C.B., Butler J., Endrizzi M.,

RA Qui D., Ianakiev P., Pedersen D., Nelson M., Washburne M.,

RA Selitrennikoff C.P., Kinsey J.A., Braun E.L., Zelter A., Schulte U.,

RA Kothe G.O., Jedd G., Mewes W., Staben C., Marcotte E., Greenberg D.,

RA Roy A., Foley K., Naylor J., Thomann N., Barrett R., Gnerre S.,

RA Kamal M., Kamvysselis M., Mauceli E., Bielke C., Rudd S., Frishman D.,

RA Krystofova S., Rasmussen C., Metzenberg R.L., Perkins D.D., Kroken S.,  
 RA Cogoni C., Macino G., Catcheside D., Li W., Pratt R.J., Osmani S.A.,  
 RA DeSouza C.C., Glass L., Orbach M.J., Berglund J., Voelker R.,  
 RA Yarden O., Plamann M., Seiler S., Dunlap J., Radford A., Aramayo R.,  
 RA Natvig D.O., Alex L.A., Mannhaupt G., Ebbbole D.J., Freitag M.,  
 RA Paulsen I., Sachs M.S., Lander E.S., Nusbaum C., Birren B.;  
 RT "The Genome Sequence of the Filamentous Fungus *Neurospora crassa*.";   
 RL Nature 0:0-0(2003).  
 CC -!- SIMILARITY: Belongs to the Ser/Thr protein kinase family.  
 CC -!- CAUTION: The sequence shown here is derived from an  
 CC EMBL/GenBank/DDBJ whole genome shotgun (WGS) entry which is  
 CC preliminary data.  
 DR EMBL; AABX01000374; EAA30285.1; -.  
 DR GO; GO:0005524; F:ATP binding; IEA.  
 DR GO; GO:0004674; F:protein serine/threonine kinase activity; IEA.  
 DR GO; GO:0016740; F:transferase activity; IEA.  
 DR GO; GO:0006468; P:protein amino acid phosphorylation; IEA.  
 DR InterPro; IPR000719; Prot\_kinase.  
 DR InterPro; IPR010513; Ribonuc\_2-5A.  
 DR InterPro; IPR008271; Ser\_thr\_pkin\_AS.  
 DR Pfam; PF00069; Pkinase; 1.  
 DR Pfam; PF06479; Ribonuc\_2-5A; 1.  
 DR ProDom; PD000001; Prot\_kinase; 1.  
 DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; UNKNOWN\_1.  
 DR PROSITE; PS50011; PROTEIN\_KINASE\_DOM; 1.  
 DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
 KW ATP-binding; Hypothetical protein; Kinase;  
 KW Serine/threonine-protein kinase; Transferase.  
 SQ SEQUENCE 1208 AA; 133859 MW; 877E98F8186E6AA8 CRC64;

Query Match 57.5%; Score 46; DB 2; Length 1208;  
 Best Local Similarity 66.7%; Pred. No. 77;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 3 DGNFFGAGLDHQ 14  
 || ||| ||| :  
 Db 375 DGTFFGFGLDRE 386

## RESULT 2

### Q8IAN1

ID Q8IAN1 PRELIMINARY; PRT; 1245 AA.  
 AC Q8IAN1;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last annotation update)  
 DE Hypothetical protein PF08\_0127.  
 GN Name=PF08\_0127;  
 OS Plasmodium falciparum (isolate 3D7).  
 OC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RA Seeger K., Murphy L., Harris D., Berriman M., Pain A., Hall N.,  
 RA Quail M., Barrell B.;  
 RL Submitted (SEP-2002) to the EMBL/GenBank/DDBJ databases.  
 DR EMBL; AL844507; CAD51332.1; -.

KW Hypothetical protein.

SQ SEQUENCE 1245 AA; 147911 MW; D856486AFDFE4DDF CRC64;

Query Match 56.2%; Score 45; DB 2; Length 1245;

Best Local Similarity 72.7%; Pred. No. 1.2e+02;

Matches 8; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGL 11

||: |||| ||

Db 305 YSNNNFFGQGL 315

### RESULT 3

Q883V9

ID Q883V9 PRELIMINARY; PRT; 1562 AA.

AC Q883V9;

DT 01-JUN-2003 (TrEMBLrel. 24, Created)

DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE YD repeat protein.

GN OrderedLocusNames=PSPTO2239;

OS Pseudomonas syringae (pv. tomato).

OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;

OC Pseudomonadaceae; Pseudomonas.

OX NCBI\_TaxID=323;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DC3000;

RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;

RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,

RA Gwinn M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,

RA Madupu R., Daugherty S.C., Brinkac L.M., Beanan M.J., Haft D.H.,

RA Nelson W.C., Davidsen T.M., Zafar N., Zhou L., Liu J., Yuan Q.,

RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,

RA Utterback T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,

RA Deng W.-L., Ramos A.R., Alfano J.R., Cartinhour S., Chatterjee A.K.,

RA Delaney T.P., Lazarowitz S.G., Martin G.B., Schneider D.J., Tang X.,

RA Bender C.L., White O., Fraser C.M., Collmer A.;

RT "The complete genome sequence of the Arabidopsis and tomato pathogen

RT Pseudomonas syringae pv. tomato DC3000.";

RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).

DR EMBL; AE016863; AAO55755.1; -.

DR TIGR; PSPTO2239; -.

DR InterPro; IPR000977; DNA\_ligase.

DR InterPro; IPR006530; YD.

DR Pfam; PF05593; RHS\_repeat; 6.

DR TIGRFAMs; TIGR01643; YD\_repeat\_2x; 3.

DR PROSITE; PS00697; DNA\_LIGASE\_A1; UNKNOWN\_1.

KW Complete proteome.

SQ SEQUENCE 1562 AA; 175713 MW; 8DC10DA1BFE37BF1 CRC64;

Query Match 55.6%; Score 44.5; DB 2; Length 1562;

Best Local Similarity 58.8%; Pred. No. 1.8e+02;

Matches 10; Conservative 0; Mismatches 2; Indels 5; Gaps 1;

Qy 1 YSDG-----NFFGAGLD 12

| || || ||||

Db 295 YKDGAGRERNFLGAGLD 311

RESULT 4

Q89ZY7

ID Q89ZY7 PRELIMINARY; PRT; 189 AA.  
AC Q89ZY7;  
DT 01-JUN-2003 (TrEMBLrel. 24, Created)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)  
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)  
DE Hypothetical protein.  
GN OrderedLocusNames=BT4234;  
OS Bacteroides thetaiotaomicron.  
OC Bacteria; Bacteroidetes; Bacteroides (class); Bacteroidales;  
OC Bacteroidaceae; Bacteroides.  
OX NCBI\_TaxID=818;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=VPI-5482 / ATCC 29148;  
RX MEDLINE=22550858; PubMed=12663928;  
RA Xu J., Bjursell M.K., Himrod J., Deng S., Carmichael L.K.,  
RA Chiang H.C., Hooper L.V., Gordon J.I.;  
RT "A genomic view of the human-Bacteroides thetaiotaomicron symbiosis.";  
RL Science 299:2074-2076(2003).  
DR EMBL; AE016944; AAO79339.1; -.  
KW Complete proteome; Hypothetical protein.  
SQ SEQUENCE 189 AA; 21766 MW; 4BC051EA64A7BC20 CRC64;

Query Match 55.0%; Score 44; DB 2; Length 189;  
Best Local Similarity 70.0%; Pred. No. 27;  
Matches 7; Conservative 2; Mismatches 1; Indels 0; Gaps 0;

Qy 4 GNFFGAGLDH 13  
||| ||| : :  
Db 122 GNFFGAGISY 131

RESULT 5

KC2B\_CAEEL

ID KC2B\_CAEEL STANDARD; PRT; 234 AA.  
AC P28548; O62352; Q22077;  
DT 01-DEC-1992 (Rel. 24, Created)  
DT 01-DEC-1992 (Rel. 24, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Casein kinase II beta chain (CK II).  
GN Name=kin-5; Synonyms=kin-10; ORFNames=T01G9.6;  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;  
RN [1]  
RP SEQUENCE FROM N.A. (ISOFORM A).  
RX MEDLINE=92011787; PubMed=1918084;  
RA Hu E., Rubin C.S.;  
RT "Casein kinase II from Caenorhabditis elegans. Cloning,  
RT characterization, and developmental regulation of the gene encoding  
RT the beta subunit.";

RL J. Biol. Chem. 266:19796-19802(1991).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=Bristol N2;  
 RA Lennard N.;  
 RL Submitted (JUL-1996) to the EMBL/GenBank/DDBJ databases.  
 RN [3]  
 RP REVISIONS, AND ALTERNATIVE SPLICING.  
 RA Durbin R.;  
 RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- FUNCTION: Participates in Wnt signaling. Plays a complex role in  
 CC regulating the basal catalytic activity of the alpha subunit (By  
 CC similarity).  
 CC -!- SUBUNIT: Tetramer of two alpha and two beta chains.  
 CC -!- ALTERNATIVE PRODUCTS:  
 CC Event=Alternative splicing; Named isoforms=2;  
 CC Name=a;  
 CC IsoId=P28548-1; Sequence=Displayed;  
 CC Name=b;  
 CC IsoId=P28548-2; Sequence=VSP\_001093;  
 CC Note=No experimental confirmation available;  
 CC -!- DEVELOPMENTAL STAGE: Elevated levels are observed during  
 CC embryogenesis, liver regeneration, and adipocyte differentiation.  
 CC -!- PTM: Phosphorylated by alpha chain (By similarity).  
 CC -!- SIMILARITY: Belongs to the casein kinase 2 beta chain family.  
 CC -----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC -----  
 DR EMBL; M73827; AAA27983.1; -.  
 DR EMBL; Z75713; CAB00056.1; -.  
 DR EMBL; Z75713; CAB00053.1; -.  
 DR PIR; T24317; T24317.  
 DR HSSP; P13862; 1QF8.  
 DR IntAct; P28548; -.  
 DR WormPep; T01G9.6a; CE18168.  
 DR WormPep; T01G9.6b; CE06343.  
 DR InterPro; IPR000704; CAS\_kinase\_II.  
 DR Pfam; PF01214; CK\_II\_beta; 1.  
 DR PRINTS; PR00472; CASNKINASEII.  
 DR ProDom; PD003829; CAS\_kinase\_II; 1.  
 DR PROSITE; PS01101; CK2\_BETA; 1.  
 KW Alternative splicing; Phosphorylation;  
 KW Serine/threonine-protein kinase; Transferase; Wnt signaling pathway.  
 FT MOD\_RES 2 2 Phosphoserine (by autocatalysis)  
 FT (Probable).  
 FT DOMAIN 55 63 Asp/Glu-rich (acidic).  
 FT VARSPLIC 58 58 P -> PE (in isoform b).  
 FT /FTId=VSP\_001093.  
 FT CONFLICT 141 141 M -> D (in Ref. 2).  
 SQ SEQUENCE 234 AA; 26452 MW; A0814A48B768347D CRC64;



Query Match 55.0%; Score 44; DB 1; Length 234;  
 Best Local Similarity 46.2%; Pred. No. 34;  
 Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLDH 13  
 ::||::|| | |  
 Db 152 HTDGSYFGTGFPH 164

# RESULT 6

## Q7VA01

ID Q7VA01 PRELIMINARY; PRT; 1524 AA.  
 AC Q7VA01;  
 DT 01-OCT-2003 (TrEMBLrel. 25, Created)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Ferredoxin-dependent glutamate synthase (EC 1.4.7.1).  
 GN Name=glfF; OrderedLocusNames=Pro1668;  
 OS Prochlorococcus marinus.  
 OC Bacteria; Cyanobacteria; Prochlorophytes; Prochlorococcaceae;  
 OC Prochlorococcus.  
 OX NCBI\_TaxID=1219;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=SARG / CCMP 1375 / SS120;  
 RX MEDLINE=22810154; PubMed=12917486;  
 RA Dufresne A., Salanoubat M., Partensky F., Artiguenave F., Axmann I.M.,  
 RA Barbe V., Duprat S., Galperin M.Y., Koonin E.V., Le Gall F.,  
 RA Makarova K.S., Ostrowski M., Oztas S., Robert C., Rogozin I.B.,  
 RA Scanlan D.J., Tandeau de Marsac N., Weissenbach J., Wincker P.,  
 RA Wolf Y.I., Hess W.R.;  
 PT "Genome sequence of the cyanobacterium Prochlorococcus marinus SS120,  
 RT a nearly minimal oxyphototrophic genome.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10020-10025(2003).  
 DR EMBL; AE017166; AAQ00712.1; -.  
 DR GO; GO:0016041; F:glutamate synthase (ferredoxin) activity; IEA.  
 DR GO; GO:0016491; F:oxidoreductase activity; IEA.  
 DR GO; GO:0006537; P:glutamate biosynthesis; IEA.  
 DR GO; GO:0006807; P:nitrogen metabolism; IEA.  
 DR InterPro; IPR003009; FMN\_enzyme.  
 DR InterPro; IPR002932; Glu\_synthase.  
 DR InterPro; IPR002489; Glu\_synthase\_C.  
 DR InterPro; IPR006982; Glu\_synth\_cent.  
 DR InterPro; IPR006981; Glu\_synth\_NTN.  
 DR Pfam; PF01645; Glu\_synthase; 1.  
 DR Pfam; PF04897; Glu\_synth\_NTN; 1.  
 DR Pfam; PF04898; Glu\_syn\_central; 1.  
 DR Pfam; PF01493; GXGXG; 1.  
 KW Complete proteome; Oxidoreductase.  
 SQ SEQUENCE 1524 AA; 167372 MW; F8FDFD7FB3D2C92B CRC64;

Query Match 55.0%; Score 44; DB 2; Length 1524;  
 Best Local Similarity 66.7%; Pred. No. 2.1e+02;  
 Matches 8; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLD 12  
 :|||:| || ||

## RESULT 7

## GLTS\_SYNY3

ID GLTS\_SYNY3 STANDARD; PRT; 1556 AA.

AC P55038; Q59980;

DT 01-OCT-1996 (Rel. 34, Created)

DT 01-NOV-1997 (Rel. 35, Last sequence update)

DT 01-OCT-2004 (Rel. 45, Last annotation update)

DE Ferredoxin-dependent glutamate synthase 2 (EC 1.4.7.1) (FD-GOGAT).

GN Name=gltS; OrderedLocusNames=sll1499;

OS Synechocystis sp. (strain PCC 6803).

OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.

OX NCBI\_TaxID=1148;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=95244836; PubMed=7727752;

RA Navarro F., Chavez S., Candau P., Florencio F.J.;

RT "Existence of two ferredoxin-glutamate synthases in the cyanobacterium

RT Synechocystis sp. PCC 6803. Isolation and insertional inactivation of

RT gltB and gltS genes.";

RL Plant Mol. Biol. 27:753-767(1995).

RN [2]

RP SEQUENCE FROM N.A.

RA Terauchi K., Ikeuchi M., Ohmori M.;

RL Submitted (DEC-1995) to the EMBL/GenBank/DDBJ databases.

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=97061201; PubMed=8905231;

RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,

RA Miyajima N., Hirose M., Sugiura M., Sasamoto S., Kimura T.,

RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,

RA Shimpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,

RA Tabata S.;

RT "Sequence analysis of the genome of the unicellular cyanobacterium

RT Synechocystis sp. strain PCC6803. II. Sequence determination of the

RT entire genome and assignment of potential protein-coding regions.";

RL DNA Res. 3:109-136(1996).

CC -!- CATALYTIC ACTIVITY: 2 L-glutamate + 2 oxidized ferredoxin = L-  
CC glutamine + 2-oxoglutarate + 2 reduced ferredoxin.

CC -!- COFACTOR: Binds a 3Fe-4S cluster; FAD and FMN.

CC -!- PATHWAY: Glutamine synthetase/GOGAT pathway which is involved in  
CC the assimilation of ammonia.

CC -!- SIMILARITY: Belongs to the glutamate synthase family.

CC

CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC

DR EMBL; X92480; CAA63218.1; -.

DR EMBL; D78371; BAA11379.1; -.

DR EMBL; D90916; BAA18693.1; -.

DR PIR; S76781; S76781.  
 DR PDB; 1LLW; X-ray; A=37-1556.  
 DR PDB; 1LLZ; X-ray; A=37-1556.  
 DR PDB; 1LM1; X-ray; A=37-1556.  
 DR PDB; 1OFD; X-ray; A/B=37-1556.  
 DR PDB; 1OFE; X-ray; A/B=37-1556.  
 DR InterPro; IPR003009; FMN\_enzyme.  
 DR InterPro; IPR002932; Glu\_synthase.  
 DR InterPro; IPR002489; Glu\_synthase\_C.  
 DR InterPro; IPR006982; Glu\_synth\_cent.  
 DR InterPro; IPR006981; Glu\_synth\_NTN.  
 DR Pfam; PF01645; Glu\_synthase; 1.  
 DR Pfam; PF04897; Glu\_synth\_NTN; 1.  
 DR Pfam; PF04898; Glu\_syn\_central; 1.  
 DR Pfam; PF01493; GXGXG; 1.  
 KW 3D-structure; 3Fe-4S; Complete proteome; FAD; Flavoprotein; FMN;  
 KW Glutamate biosynthesis; Iron-sulfur; Oxidoreductase.  
 FT DOMAIN 37 384 Glutamine amidotransferase (Potential).  
 FT METAL 1173 1173 Iron-sulfur (3Fe-4S) (By similarity).  
 FT METAL 1179 1179 Iron-sulfur (3Fe-4S) (By similarity).  
 FT METAL 1184 1184 Iron-sulfur (3Fe-4S) (By similarity).  
 FT CONFLICT 491 491 E -> Q (in Ref. 1).  
 FT CONFLICT 570 572 ESA -> NPR (in Ref. 1).  
 FT CONFLICT 642 650 GAILTENQS -> RRNIGLRIKV (in Ref. 1).  
 FT CONFLICT 659 659 G -> E (in Ref. 1).  
 FT CONFLICT 940 941 GG -> PP (in Ref. 1).  
 FT CONFLICT 1059 1059 H -> L (in Ref. 1).  
 FT CONFLICT 1295 1295 E -> RK (in Ref. 1).  
 FT CONFLICT 1310 1310 V -> D (in Ref. 1).  
 FT CONFLICT 1323 1323 A -> S (in Ref. 1).  
 FT CONFLICT 1531 1531 Missing (in Ref. 1).  
 SQ SEQUENCE 1556 AA; 169498 MW; 4BDAD5F9A4064D9D CRC64;  
  
 Query Match 55.0%; Score 44; DB 1; Length 1556;  
 Best Local Similarity 66.7%; Pred. No. 2.2e+02;  
 Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 YSDGNFFGAGLD 12  
 :||| ||||  
 Db 370 FSDGKIVGAGLD 381

# RESULT 8

Q983H6

ID Q983H6 PRELIMINARY; PRT; 2863 AA.  
 AC Q983H6;  
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)  
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Cyclic beta 1-2 glucan synthetase.  
 GN OrderedLocusNames=mlr8325;  
 OS Rhizobium loti (Mesorhizobium loti).  
 CC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;  
 OC Phyllobacteriaceae; Mesorhizobium.  
 OX NCBI\_TaxID=381;  
 RN [1]  
 RP SEQUENCE FROM N.A.

RC STRAIN=MAFF303099;  
 RX MEDLINE=21082930; PubMed=11214968;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti.";  
 RL DNA Res. 7:331-338(2000).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=MAFF303099;  
 RX MEDLINE=21082936; PubMed=11214974;  
 RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,  
 RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,  
 RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,  
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,  
 RA Takeuchi C., Yamada M., Tabata S.;  
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium  
 RT Mesorhizobium loti (supplement).";  
 RL DNA Res. 7:381-406(2000).  
 DR EMBL; AP003013; BAB53905.1; -.  
 DR InterPro; IPR009342; CBM\_X.  
 DR InterPro; IPR010383; Glyco\_transf\_36.  
 DR InterPro; IPR008928; Glyco\_trans\_6hp.  
 DR InterPro; IPR010403; GT36\_AF.  
 DR Pfam; PF06204; CBM\_X; 2.  
 DR Pfam; PF06165; Glyco\_transf\_36; 2.  
 DR Pfam; PF06205; GT36\_AF; 2.  
 KW Complete proteome.  
 SQ SEQUENCE 2863 AA; 316778 MW; F160E5DDB74E0246 CRC64;

Query Match 55.0%; Score 44; DB 2; Length 2863;  
 Best Local Similarity 61.5%; Pred. No. 3.9e+02;  
 Matches 8; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
 :|||:| |||  
 Db 722 FSDGSFTGKGLYH 734

# RESULT 9

Q8IIW5

ID Q8IIW5 PRELIMINARY; PRT; 245 AA.  
 AC Q8IIW5;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)  
 DE Casein kinase II beta chain, putative.  
 GN ORFNames=PF11\_0048;  
 OS Plasmodium falciparum (isolate 3D7).  
 CC Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.  
 OX NCBI\_TaxID=36329;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=22255705; PubMed=12368864;

RA Gardner M.J., Hall N., Fung E., White O., Berriman M., Hyman R.W.,  
 RA Carlton J.M., Pain A., Nelson K.E., Bowman S., Paulsen I.T., James K.,  
 RA Eisen J.A., Rutherford K., Salzberg S.L., Craig A., Kyes S.,  
 RA Chan M.S., Nene V., Shallom S.J., Suh B., Peterson J., Angiuoli S.,  
 RA Pertea M., Allen J., Selengut J., Haft D., Mather M.W., Vaidya A.B.,  
 RA Martin D.M., Fairlamb A.H., Fraunholz M.J., Roos D.S., Ralph S.A.,  
 RA McFadden G.I., Cummings L.M., Subramanian G.M., Mungall C.,  
 RA Venter J.C., Carucci D.J., Hoffman S.L., Newbold C., Davis R.W.,  
 RA Fraser C.M., Barrell B.;  
 RT "Genome sequence of the human malaria parasite Plasmodium  
 RT falciparum.";  
 RL Nature 419:498-511(2002).  
 DR EMBL; AE014836; AAN35637.1; -.  
 DR HSSP; P13862; 1QF8.  
 DR GO; GO:0005956; C:protein kinase CK2 complex; IEA.  
 DR GO; GO:0016301; F:kinase activity; IEA.  
 DR GO; GO:0008605; F:protein kinase CK2 regulator activity; IEA.  
 DR InterPro; IPR000704; CAS\_kinase\_II.  
 DR Pfam; PF01214; CK\_II\_beta; 1.  
 DR PRINTS; PR00472; CASNKINASEII.  
 DR ProDom; PD003829; CAS\_kinase\_II; 1.  
 KW Kinase.  
 SQ SEQUENCE 245 AA; 28365 MW; BD895A4C34124E67 CRC64;

Query Match 53.8%; Score 43; DB 2; Length 245;  
 Best Local Similarity 53.8%; Pred. No. 53;  
 Matches 7; Cconservative 1; Mismatches 5; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
 | | | : | | | |  
 Db 161 YLDGSFFGTSFPH 173

# RESULT 10

## CPK4\_ONCMY

ID CPK4\_ONCMY STANDARD; PRT; 504 AA.  
 AC O93297;  
 DT 15-JUL-1999 (Rel. 38, Created)  
 DT 15-JUL-1999 (Rel. 38, Last sequence update)  
 DT 05-JUL-2004 (Rel. 44, Last annotation update)  
 DE Cytochrome P450 2K4 (EC 1.14.14.1) (CYPIIK4).  
 GN Name=CYP2K4;  
 OS Oncorhynchus mykiss (Rainbow trout) (Salmo gairdneri).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Actinopterygii; Neopterygii; Teleostei; Euteleostei;  
 OC Protacanthopterygii; Salmoniformes; Salmonidae; Oncorhynchus.  
 OX NCBI\_TaxID=8022;  
 RN [1]  
 RF SEQUENCE FROM N.A.  
 RC TISSUE=Kidney;  
 RA Yang Y.-H., Andersson T.B., Ryu B.-W., Wang J.-L., Buhler D.R.;  
 RT "CYP2K4: a new cytochrome P450 isoform from male trunk kidney of post-  
 RT spawning rainbow trout.";  
 RL Submitted (JAN-1998) to the EMBL/GenBank/DDBJ databases.  
 CC -!- CATALYTIC ACTIVITY: RH + reduced flavoprotein + O(2) = ROH +  
 CC oxidized flavoprotein + H(2)O.  
 CC -!- SUBCELLULAR LOCATION: Membrane-bound. Endoplasmic reticulum (By

CC similarity).

CC -!- SIMILARITY: Belongs to the cytochrome P450 family.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; AF043296; AAC26492.1; -.

DR HSSP; P00179; 1DT6.

DR InterPro; IPR001128; Cytochrome\_P450.

DR InterPro; IPR002401; EP450I.

DR Pfam; PF00067; p450; 1.

DR PRINTS; PR00463; EP450I.

DR PRINTS; PR00385; P450.

DR PROSITE; PS00086; CYTOCHROME\_P450; 1.

KW Electron transport; Endoplasmic reticulum; Heme; Membrane; Microsome;

KW Monooxygenase; Oxidoreductase.

FT METAL 447 447 Iron (heme axial ligand) (By similarity).

SQ SEQUENCE 504 AA; 56734 MW; AC4292C18617C4B1 CRC64;

Query Match 53.8%; Score 43; DB 1; Length 504;

Best Local Similarity 66.7%; Pred. No. 1.1e+02;

Matches 8; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLD 12

: | | | | |

Db 302 FSIGNLFGAGTD 313

# RESULT 11

## Q7UJE2

ID Q7UJE2 PRELIMINARY; PRT; 714 AA.

AC Q7UJE2;

DT 01-OCT-2003 (TrEMBLrel. 25, Created)

DT 01-OCT-2003 (TrEMBLrel. 25, Last sequence update)

DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)

DE Probable secreted protein-putative xanthan lyase related.

GN OrderedLocusNames=RB11948;

OS Rhodopirellula baltica.

OC Bacteria; Planctomycetes; Planctomycetacia; Planctomycetales;

OC Planctomycetaceae; Pirellula.

OX NCBI\_TaxID=117;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=1;

RX MEDLINE=22735913; PubMed=12835416;

RA Gloeckner F.O., Kube M., Bauer M., Teeling H., Lombardot T.,

RA Ludwig W., Gade D., Beck A., Borzym K., Heitmann K., Rabus R.,

RA Schlesner H., Amann R., Reinhardt R.;

RT "Complete genome sequence of the marine planctomycete Pirellula sp.

RT strain 1.";

RL Proc. Natl. Acad. Sci. U.S.A. 100:8298-8303(2003).

DR EMBL; BX294154; CAD77316.1; -.

DR GO; GO:0016829; F:lyase activity; IEA.  
KW Complete proteome; Lyase.  
SQ SEQUENCE 714 AA; 79349 MW; E3565E862778F0F9 CRC64;

Query Match 53.8%; Score 43; DB 2; Length 714;  
Best Local Similarity 46.2%; Pred. No. 1.5e+02;  
Matches 6; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

QY 1 YSDGNFFGAGLDH 13  
:: |||:|:| |  
Db 594 HASGNFYGSGYHH 606

# RESULT 12

## YAGX\_ECOLI

ID YAGX\_ECOLI STANDARD; PRT; 841 AA.  
AC P77802;  
DT 01-NOV-1997 (Rel. 35, Created)  
DT 01-NOV-1997 (Rel. 35, Last sequence update)  
DT 05-JUL-2004 (Rel. 44, Last annotation update)  
DE Hypothetical protein yagX precursor.  
GN Name=yagX; OrderedLocusNames=b0291;  
OS Escherichia coli.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
OC Enterobacteriaceae; Escherichia.  
OX NCBI\_TaxID=562;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=K12 / MG1655;  
RX MEDLINE=97426617; PubMed=9278503;  
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,  
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,  
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,  
RA Mau B., Shao Y.;  
RT "The complete genome sequence of Escherichia coli K-12.";  
RL Science 277:1453-1474 (1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,  
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H., Lin D.,  
RA Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;  
RL Submitted (NOV-1996) to the EMBL/GenBank/DDBJ databases.  
CC -!- SIMILARITY: Some, to E.coli plasmid NTP513 CFA fimbria subunit C  
CC (cfaC).  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; AE000136; AAC73394.1; -.  
DR EMBL; U73857; AAB18020.1; -.  
DR PIR; C64755; C64755.  
DR EchoBASE; EB3333; -.

DR EcoGene; EG13563; yagX.  
 DR InterPro; IPR000627; Dioxxygenase.  
 KW Complete proteome; Hypothetical protein; Signal.  
 FT SIGNAL 1 29 Potential.  
 FT CHAIN 30 841 Hypothetical protein yagX.  
 SQ SEQUENCE 841 AA; 91228 MW; D2016BB0ACD726AC CRC64;

Query Match 53.8%; Score 43; DB 1; Length 841;  
 Best Local Similarity 63.6%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNFFGAGLDHQ 14  
 ||:| ||: ||  
 Db 562 GNWFSAGMTHQ 572

# RESULT 13

## Q7AHC2

ID Q7AHC2 PRELIMINARY; PRT; 841 AA.  
 AC Q7AHC2;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Putative enzyme.  
 GN OrderedLocusNames=ECs0321;  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 CC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / RIMD 0509952 / EHEC;  
 RX MEDLINE=21156231; PubMed=11258796;  
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,  
 RA Han C.-G., Ohtsubo E., Nakayama K., Murata T., Tanaka M., Tobe T.,  
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,  
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;  
 RT "Complete genome sequence of enterohemorrhagic Escherichia coli  
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";  
 RL DNA Res. 8:11-22(2001).  
 DR EMBL; AP002551; BAB33744.1; -.  
 DR InterPro; IPR000627; Dioxxygenase.  
 DR InterPro; IPR000577; FGGY\_kin.  
 DR PROSITE; PS00445; FGGY\_KINASES\_2; UNKNOWN\_1.  
 SQ SEQUENCE 841 AA; 91227 MW; DCCR0ACA1CB821E5 CRC64;

Query Match 53.8%; Score 43; DB 2; Length 841;  
 Best Local Similarity 63.6%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNFFGAGLDHQ 14  
 ||:| ||: ||  
 Db 562 GNWFSAGMTHQ 572

# RESULT 14

## Q8CWC1



ID Q8CWC1 PRELIMINARY; PRT; 841 AA.  
 AC Q8CWC1;  
 DT 01-MAR-2003 (TrEMBLrel. 23, Created)  
 DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Hypothetical protein yagX.  
 GN Name=yagX; OrderedLocusNames=c0402;  
 OS Escherichia coli O6.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=217992;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O6:H1 / CFT073 / ATCC 700928;  
 RX MEDLINE=22388234; PubMed=12471157;  
 RA Welch R.A., Burland V., Plunkett G. III, Redford P., Roesch P.,  
 RA Rasko D., Buckles E.L., Liou S.-R., Boutin A., Hackett J., Stroud D.,  
 RA Mayhew G.F., Rose D.J., Zhou S., Schwartz D.C., Perna N.T.,  
 RA Mobley H.L.T., Donnenberg M.S., Blattner F.R.;  
 RT "Extensive mosaic structure revealed by the complete genome sequence  
 RT of uropathogenic Escherichia coli.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 99:17020-17024 (2002).  
 DR EMBL; AE016756; AAN78883.1; -.  
 DR GO; GO:0003824; F:catalytic activity; IEA.  
 DR GO; GO:0008199; F:ferric iron binding; IEA.  
 DR GO; GO:0006725; P:aromatic compound metabolism; IEA.  
 DE InterPro; IPR000627; Dioxxygenase.  
 DR InterPro; IPR000577; FGGY\_kin.  
 DR PROSITE; PS00445; FGGY\_KINASES\_2; UNKNOWN\_1.  
 KW Complete proteome; Hypothetical protein.  
 SQ SEQUENCE 841 AA; 91262 MW; A92E97BA8448F713 CRC64;

Query Match 53.8%; Score 43; DB 2; Length 841;  
 Best Local Similarity 63.6%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNFFGAGLDHQ 14  
 ||:| ||: ||  
 Db 562 GNWFSAGMTHQ 572

# RESULT 15

Q8X6I4

ID Q8X6I4 PRELIMINARY; PRT; 841 AA.  
 AC Q8X6I4;  
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)  
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)  
 DT 01-MAR-2004 (TrEMBLrel. 26, Last annotation update)  
 DE Putative enzyme.  
 GN Name=yagX; OrderedLocusNames=z0358;  
 OS Escherichia coli O157:H7.  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;  
 OC Enterobacteriaceae; Escherichia.  
 OX NCBI\_TaxID=83334;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN=O157:H7 / EDL933 / ATCC 700927 / EHEC;

RX MEDLINE=21074935; PubMed=11206551;  
 RA Perna N.T., Plunkett G. III, Burland V., Mau B., Glasner J.D.,  
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,  
 RA Posfai G., Hackett J., Klink S., Boutin A., Shao Y., Miller L.,  
 RA Grotbeck E.J., Davis N.W., Lim A., Dimalanta E.T., Potamousis K.,  
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,  
 RA Welch R.A., Blattner F.R.;  
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7.";  
 RL Nature 409:529-533(2001).  
 DR EMBL; AE005206; AAG54616.1; -.  
 DR PIR; A90669; A90669.  
 DR PIR; D85519; D85519.  
 DR GO; GO:0003824; F:catalytic activity; IEA.  
 DR GO; GO:0008199; F:ferric iron binding; IEA.  
 DR GO; GO:0006725; P:aromatic compound metabolism; IEA.  
 DR InterPro; IPR000627; Dioxxygenase.  
 DR InterPro; IPR000577; FGGY\_kin.  
 DR PROSITE; PS00445; FGGY\_KINASES\_2; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 841 AA; 91227 MW; DCCB0ACA1CB821E5 CRC64;

Query Match 53.8%; Score 43; DB 2; Length 841;  
 Best Local Similarity 63.6%; Pred. No. 1.8e+02;  
 Matches 7; Conservative 2; Mismatches 2; Indels 0; Gaps 0;

QY 4 GNFFGAGLDHQ 14  
 ||:| ||: ||  
 Db 562 GNWFSAGMTHQ 572

Search completed: January 31, 2005, 13:22:45  
 Job time : 108.591 secs